Measurement of the top quark pair production cross section in proton-proton collisions at center-of-mass energies of 7 TeV in final states with a τ lepton with the ATLAS detector

María Teresa Pérez García-Estañ

Dirigido por:
Susana Cabrera Urbán

Tesis Doctoral
Junio 2014

VNIVERSITAT DE VALÈNCIA

Facultat de Física
Departament de Física Atòmica Molecular i Nuclear

Dra. Susana Cabrera Urbán
Científica Titular de Organismos Públicos de Investigación

CERTIFICA:

Que la presente memoria, '***Measurement of the top quark pair production cross section in proton-proton collisions at center-of-mass energies of*** 7 TeV ***in final states with a*** τ ***lepton with the ATLAS detector***', ha sido realizada bajo mi dirección en el *Departament de Física Atòmica, Molecular i Nuclear* de la *Universitat de València* por María Teresa Pérez García-Estañ y constituye su tesis para optar al grado de doctora en Física.

Y para que así conste, en cumplimiento de la legislación vigente, firmo el presente Certificado.

Firmado
Dra. Susana Cabrera Urbán

El trabajo descrito en esta tesis se ha llevado a cabo en el *Instituto de Física Corpuscular* (IFIC) en Valencia, España.

El IFIC es un centro mixto de la *Universitat de València* y el *Consejo Superior de Investigaciones Científicas* (CSIC)

A mis padres

Preface

The Standard Model (SM) of Particle Physics is a quantum field theory developed in the 1960's to explain the behaviour of elementary particles and the fundamental forces that govern their interactions. It is experimentally well-tested and has successfully explained a host of experimental results and precisely predicted a wide variety of phenomena. However, some experimental evidences are not explained by the Standard Model, like dark matter, the baryon asymmetry or neutrino masses.

To discover new Physics, the Large Hadron Collider (LHC), a particle accelerator and collider, was built by the European Organization for Nuclear Research (CERN). The accelerator, the detectors and the computing infrastructure to produce and study proton-proton collisions makes of the LHC one of the largest and most complex experimental facilities ever built.

The top quark physics program is of great importance at the LHC, since the top quark has the largest coupling to the Standard Model Higgs boson, responsible of the mass of fundamental particles. The high energies achieved and the large amount of data taken make high precision measurements of the top quark properties possible at the LHC, thus improving our understanding of Particle Physics and opening the door to new discoveries. In particular, the studies presented in this thesis focus on top quark pair production in the dilepton channel with an hadronically decaying tau lepton in the final state.

An introduction to the Standard Model of Particle Physics and the current situation of top quark measurements is done in Chapter 1, where the role of tau lepton in top quark physics, both for Standard Model measurements and new physic searches, is highlighted.

The data analysed have been collected by the ATLAS experiment, one of the particle detectors present at the LHC. A description of the LHC accelerator ring and the ATLAS detector is given in Chapter 2.

The simulated and real data used in the analysis and the physics objects definitions (electrons, muons, jets, missing transverse energy and τ-leptons) are detailed in Chapter 3.

On Chapter 4, τ lepton reconstruction and identification methods are discussed. Taus are identified using two methods aiming to distinguish hadronically decaying tau leptons from jets originated by quarks or gluons. Both methods are applied to early data collected with the ATLAS detector and a first assessment of the backgrounds for the $t\bar{t}$ lepton plus tau channel was made using data driven techniques.

The top pair production cross section measurement technique is developed in Chapter 5 using 2.05 fb^{-1} of data taken by the ATLAS detector in 2011. This measurement takes advantage of a Boosted Decision Tree discriminator trained to separate the tau signal from the busy jet background. Finally, some concluding remarks are given in Chapter 6.

Contents

List of Figures

List of Tables

1 Theoretical Motivations

1.1 The Standard Model of Particle Physics

The Standard Model [1] is the currently accepted and experimentally well-tested theory of fundamental particles and their interactions. It describes the Universe in terms of matter, made of fundamental particles called fermions (half-integer spin particles), and the fundamental forces that govern the interactions between these particles, which are transmitted by another set of particles called bosons (integer spin particles). It provides a unified picture of three fundamental forces: Electromagnetism, the Weak force, and the Strong force. In addition, it provides an explanation for how particle masses are created through the Higgs mechanism.

1.1.1 Fundamental forces

The interactions between fundamental particles are described by Quantum Field Theories (QFT) that combine Quantum Mechanics and Special Relativity. In these quantum field theories, particles are described as fields, ψ, that are solutions of the Euler-Lagrange equations and the interactions between these fields are mediated by gauge bosons. QFT are local gauge theories, where the Lagrangian is kept to be invariant under a local gauge transformation [2].

The Electromagnetic force is described by the Quantum Electrodynamics theory (QED), which is a local invariant gauge theory under U(1) rotations. It has infinite range and it is mediated by the photon γ, a massless particle with no electrical charge. All particles with an electric charge interact electromagnetically.

The Weak force interacts at short range, to about 10^{-18} m. It is much stronger than gravity but it is indeed the weakest of the other three. It is mediated by 3 bosons: $W^{\pm}$ and Z^0. All particles may undergo weak interactions. In the 1960's, Glashow, Salam and Weinberg proposed the Electroweak (EW) Theory as a unified description of the electromagnetic and weak interactions [3–5], invariant under a SU(2) x U(1) local gauge symmetry.

The Strong force interacts at short range (within the radius of a nucleon) and is mediated by gluons. Strong interactions are described by the theory of Quantum Chromodynamics (QCD) [6], based in the SU(3) symmetry. In this theory, particles interacting strongly possess an internal charge called "color" and gauge transformations are local transformations between particles of different colors. The force mediators (gluons) also have color charge and, therefore, they can interact strongly with other gluons. This is the strongest of all fundamental interactions and it is the responsible of binding protons and neutrons together to form atoms and also confines quarks together into hadrons. When two quarks are close to one another, they exchange gluons and create a very strong color force field that binds the quarks together. This field does not diminish when the distance between the quarks increases and therefore isolated quarks can not exist. On the other hand, the interaction between quarks gets weaker as quarks get closer together and quarks act like free particles. This behaviour is known as "asymptotic freedom".

The combination of the EW and QCD theories describe 3 out of 4 fundamental forces of Nature and constitute the Standard Model. There is a fourth fundamental force: Gravity, of infinite range but very weak compared to the other three forces. Gravity does not play a role in Elementary Particle Physics and it is not incorporated in the Standard Model.

	Gravity	Weak Force	Electromagnetic Force	Strong Force
Acts on	Mass-Energy	Flavor	Electric charge	Color charge
Mediating particles	Graviton	$Z^0, W^{\pm}$	Photons	Gluons
Experiencing particles	all	Quarks, Leptons	Electrically charged	Quarks, gluons
Strength at 10^{-18}	10^{-41}	0.8	1	25
Strength at $3 \cdot 10^{-17}$	10^{-41}	10^{-4}	1	60

Table 1.1: *Properties of the fundamental forces. The strength of the interactions is shown relative to the strength of the electromagnetic force for two u quarks separated by the specified distances.*

Table 1.1 summarizes the fundamental interactions and their properties. The strength of the interactions is displayed comparatively with the electromagnetic force for two u quarks in the distances 10^{-18} m (upper limit for the size of quarks and electrons) and $3 \cdot 10^{-17}$ m (over the range of the weak force) [7]. Quarks and other fundamental particles are described in Section 1.1.2.

In Table 1.2 the gauge bosons that mediate the fundamental interactions are shown along with their mass and other properties. Bosons obey Bose-Einstein statistics.

Force	Gauge Boson	Mass [GeV]	Charge [e]	Color charge	Spin
Strong	gluon (g)	0	0	yes	1
Electromagnetic	photon (γ)	0	0	no	1
Weak	Z^0	91.18	0	no	1
	$W^{\pm}$	80.4	± 1	no	1
Gravity	Graviton	$< 7 \cdot 10^{-32}$ eV	0	no	2

Table 1.2: *The fundamental forces with their corresponding gauge bosons. The masses of the gauge bosons are taken from [8]. Gravity is not included in the SM. The Graviton has not been found yet.*

1.1.2 Fundamental particles

The elementary particles that constitute matter are called fermions. Fermions are characterized by Fermi-Dirac statistics and follow the Pauli exclusion principle, which prevents two identical fermions from being in the same quantum state at the same time. For each fundamental particle there is an antiparticle that have the same mass but opposite electric charge. Fermions can be classified in two categories according to their interactions. Those with color charge are called quarks, and they interact via all three interactions: strong, electromagnetic and weak. Fermions without color charge, which do not interact with the strong force, are called leptons. Leptons can be charged or neutral. Charged leptons interact both electromagnetically and weakly. Neutral leptons interact only weakly and are referred to as neutrinos. Quarks have only been observed in combined or bound states called hadrons, while leptons can exist as free particles.

	Family	Particle	Charge [e]	Color charge	Mass [MeV]	Spin
Leptons	I	electron (e)	-1	no	0.511	1/2
		electron neutrino (ν_e)	0	no	< 2 eV	
	II	muon (μ)	-1	no	105.7	
		muon neutrino (ν_μ)	0	no	< 2 eV	
	III	tau (τ)	-1	no	1776.8	
		tau neutrino (ν_τ)	0	no	< 2 eV	
Quarks	I	up (u)	+2/3	yes	2.3	1/2
		down (d)	-1/3	yes	4.8	
	II	charm (c)	+2/3	yes	$1.275 \cdot 10^3$	
		strange (s)	-1/3	yes	95	
	III	top (t)	+2/3	yes	$173 \cdot 10^3$	
		bottom (b)	-1/3	yes	$4.18 \cdot 10^3$	

Table 1.3: *Table of Elementary Particles. The latest mass measurements are taken from [8].*

Up to now three generations of fermions have been found. Each generation or family consists of eight particles, which are related in pairs: two quarks, a lepton and its corresponding neutrino, and the related antiparticles. The lightest and most stable particles make up the first generation, whereas the heavier and less stable particles belong to the second and third generations. Table 1.3 shows the fundamental particles that belong to each generation with their mass, charge and spin number.

All stable matter in the universe is made from particles that belong to the first generation: the electron and its neutrino and the quarks up and down, which bind together and form protons and neutrons. Any heavier particles quickly decay to the next most stable level.

1.1.3 Higgs Boson

Under the SM theory the bosons that carry the interactions are massless, given that inserting a mass term in the Lagrangian would break gauge invariance. However, the $W^\pm$ and Z^0 bosons that mediate the weak interaction are massive. To address this problem, Higgs, Brout and Englert proposed in 1964 the mechanism of spontaneous breaking of the symmetry or the Higgs

mechanism [9–12]. The masses of the $W^{\pm}$ and Z^0 bosons results from the interactions with the Higgs field, which can also give mass to the rest of fundamental particles. The interactions with the Higgs field are mediated by a massive scalar boson, called the Higgs boson (H). The mass of the Higgs boson is the only unknown parameter of this theory.

On July 2012, two of the experiments at the LHC, ATLAS and CMS, announced the discovery of a new boson in the mass region around 125-126 GeV [13], [14]. Although the observed excess is consistent with the predicted SM Higgs boson, not all of the new particle's properties have been fully determined and the new boson is also compatible with extensions of the SM, with a composite particle, etc.

The Higgs mass is extracted from the $H \rightarrow \gamma\gamma/ZZ$ channels both in ATLAS and CMS. The combined mass measurement in ATLAS [15] is:

$$m_H = 125.5 \pm 0.2\ (\text{stat.})^{+0.5}_{-0.6}\ (\text{syst.})\ \text{GeV}, \tag{1.1}$$

which is consistent with the result from CMS [16]:

$$m_H = 125.2 \pm 0.4\ (\text{stat.}) \pm 0.5\ (\text{syst.})\ \text{GeV}. \tag{1.2}$$

Apart from the measured mass, further properties are studied. The Higgs is found to have positive parity and to favour the spin-0 hypothesis [17]. Coupling measurements give more than 3σ evidence for the Higgs-like boson production being through vector-boson fusion [18]. So far, all measurements are consistent with expectations for the Standard Model Higgs boson.

On October 2013 François Englert and Peter Higgs were awarded the 2013 Nobel Prize in Physics for the theoretical discovery of the Higgs mechanism which was first formulated in 1964 [9–12].

1.1.4 Standard Model extensions

Although the Standard Model has succeeded in explaining a wide variety of experimental results, it is nevertheless incomplete and fails to be a complete "Theory of Everything". It incorporates only three out of the four fundamental forces, omitting gravity, and it does not provide a unification of the strong and the electroweak forces. There are also important questions it cannot answer, such as the origin of dark matter, the matter-antimatter asymmetry, the oscillation of neutrinos, etc. For these reasons, in High Energy Physics it is believed that there must be a larger theory which describes how fundamental matter interacts, relegating the Standard Model to a limiting case or approximation of the larger theory.

The most popular extension of the Standard Model is Supersymmetry (SUSY). SUSY is a gauge theory which generalizes the space-time symmetries of quantum field theory that transforms fermions into bosons and vice-versa. As a consequence, all SM particles have supersymmetric partners with the same quantum numbers but with the spin differing by $\pm 1/2$ [19]. The Minimal Supersymmetric Standard Model (MSSM) is a minimal extension of the Standard Model that includes supersymmetry. The MSSM model predict the existence of five Higgs bosons, three neutral ones (h, H, A) and a pair of charged Higss bosons ($H^{\pm}$) [20, 21].

Whereas in the SM there is only one Higgs doublet field responsible to give masses to all particles, in supersymmetric theories two Higgs doublets are required to generates mass for both "up" and "down" type quarks and charged leptons. One of the two Higgs doublet couples to up-type fermions and the other doublet couples to down-type fermions with different vacuum

expectation values v_u and v_d, which are normalized to the Standard Model Higgs vacuum expectation value (v):

$$\sqrt{v_u^2 + v_d^2} = v = 246 \text{ GeV}. \tag{1.3}$$

The ratio

$$\tan\beta = \frac{v_u}{v_d} \tag{1.4}$$

is a free parameter of the theory. The event topology of the charged Higgs decays are characterized by the Higgs mass ($m_{H^\pm}$) and the $\tan\beta$ parameter values.

If a MSSM charged Higgs boson exists, and its mass is lower than the top quark mass minus the bottom quark mass (light charged Higgs boson), the top quark predominantly decays into a charged Higgs boson and a b-quark. For charged Higgs bosons with $m_{H^+} > m_{top}$ (heavy charged Higgs bosons), the main production mode is top quark associated (tH^+). For large values of $\tan\beta$ the charged Higgs decays predominantly to a tau lepton and a neutrino.

SUSY is necessary in some formulations of string theory and a quantum theory of gravity is compatible with it. It may also hold the explanation for matter-antimatter asymmetry and the lightest supersymmetric particle is a candidate to explain dark matter. Moreover, SUSY suggests the unification of all gauge interactions at high energies. All these reasons have made searches for supersymmetric particles an important part of the analysis programs at the LHC, although no sign of it has been found so far. See, for instance, a summary of the most representative searches for supersymmetric particles and their mass reach carried out by the ATLAS collaboration at Ref. [22].

Hence there are many compelling reasons to investigate the TeV energy scale.

1.2 Top quark physics at the LHC

What is the role of top quark physics in the investigation of the Standard Model? The top quark is the heaviest elementary particle and it was discovered in 1995 by the CDF [23] and DØ [24] experiments of the Tevatron hadron collider in Fermilab [25]. The high value of the top quark mass and its closeness to the electroweak scale makes possible that the top quark could have a special role in the electroweak symmetry breaking. Being the heaviest fermion, it has the strongest coupling to the SM Higgs field and consequently the largest Yukawa coupling ($\lambda \sim 0.995$). Therefore it is a natural candidate to look for new physics phenomena [26].

1.2.1 Top quark production

In hadron colliders there are two mechanisms of production of top quarks: top quark pairs ($t\bar{t}$) which are produced through QCD interactions, through both gluon-gluon and quark-antiquark scattering (see figure 1.1), and single tops which are produced through weak interactions.

Top quark pair production depends on the center-of-mass energy of the collision, and the nature of the beams. At the LHC, the gluon scattering process dominates ($\sim 90\%$ of the cases at a center-of-mass energy of $\sqrt{s} = 7$ TeV) while at Tevatron, with $\sqrt{s} = 1.96$ TeV, the production of top quark pairs is kinematically restricted to the quark dominated region. This difference appears also in the values of the production cross section, $\sigma(t\bar{t})$, that is about 20 times larger at

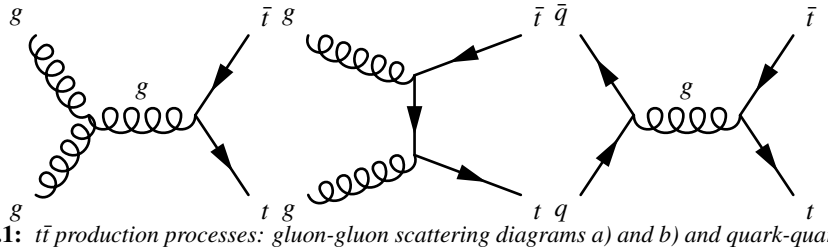


Figure 1.1: *$t\bar{t}$ production processes: gluon-gluon scattering diagrams a) and b) and quark-quark scattering diagram c).*

the LHC than at the Tevatron. The dependence of the $t\bar{t}$ production cross section with the center-of-mass energy is illustrated in Figure 1.2. At low center-of-mass energies $\sigma(t\bar{t})$ is significantly higher for proton-antiproton ($p\bar{p}$) collisions, while for proton-proton (pp) collisions it is too low to perform top quark physics studies. For larger values of $\sqrt{s}$ the gluon fusion processes dominates and $\sigma(t\bar{t})$ for both pp and $p\bar{p}$ converge.

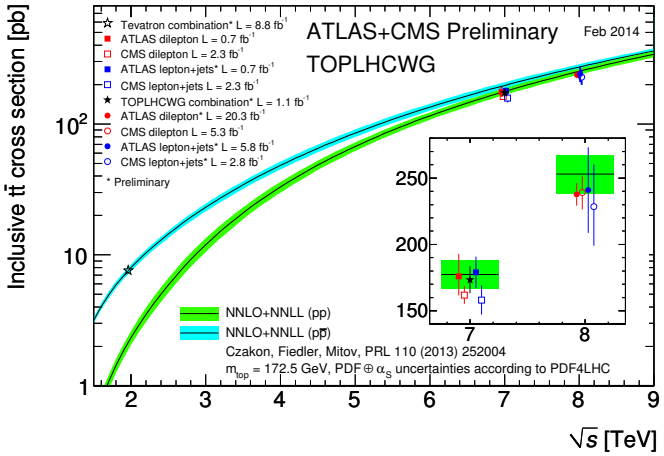


Figure 1.2: *Summary of LHC and Tevatron measurements of the top pair production cross section as a function of the centre-of-mass energy compared to the NNLO QCD calculation complemented with NNLL resummation. The theory band represents uncertainties due to renormalisation and factorisation scale, parton density functions and the strong coupling. The measurements and the theory calculation is quoted at m_{top} = 172.5 GeV.*

Results shown in this thesis consider the following theoretical $t\bar{t}$ production cross section

$$\sigma_{t\bar{t}}^{\text{theo}} = 164.57^{+11.45}_{-15.78}\ \text{pb} \tag{1.5}$$

for the generation of simulated samples and normalization of the results. The cross section calculation has been done at next-to-next-to-leading order (NNLO) in perturbative QCD for a center-of-mass energy $\sqrt{s}$ = 7 TeV [27], [28], [29]. More recent and accurate theoretical calculations at full NNLO has been done for both 7 TeV and 8 TeV center-of-mass energy. These

results are shown in Table 1.4.

σ (pb)		Reference
$\sqrt{s}$ = 7 TeV	$\sqrt{s}$ = 8 TeV	
$167.0^{+6.7}_{-10.7}$ (scale) $^{+4.6}_{-4.7}$ (pdf)	$239.1^{+9.2}_{-14.8}$ (scale) $^{+6.1}_{-6.2}$ (pdf)	[30]

Table 1.4: *Top pair production cross section at the LHC at* $\sqrt{s}$ = 7 *and* $\sqrt{s}$ = 8 TeV.

Top quarks, once produced, decay immediately, and can only be detected through their decay products. The Standard Model predicts that top quarks decay very rapidly through the weak interaction into a W boson and a b quark nearly 100% of the time. The decay modes of the W boson define the possible final states. In about $1/3$ of the cases the W boson decays into a lepton and a neutrino ($W \to \ell\nu$). All three lepton flavors (e, μ, τ) are produced at approximately equal rates. In the remaining $2/3$ of the cases it decays into a quark/anti-quark pair ($W \to q\bar{q}$).

Therefore, top quark pair events can be classify by the number of W bosons that decay leptonically, that in fact characterizes their experimental signature in the detector:

- Fully hadronic: represents about 4/9 of the $t\bar{t}$ decays. Both W bosons decay hadronically. The detector signature is very similar to strongly produced multi-jet events, which have a large cross section, thus making the hadronic channel difficult to study.
- Semileptonic: represents about 4/9 of the $t\bar{t}$ decays. Requiring one identified lepton greatly decreases the amount of QCD multi-jet background, making this channel easier to study.
- Fully leptonic: represents about 1/9 of the $t\bar{t}$ events. Both W bosons decay into a lepton-neutrino pair. It has the smallest branching fraction but, also, the smallest QCD multi-jet background. Experimentally, studies of final states containing tau leptons are generally considered separately from those containing an e or a μ.

The decay of top quark pairs, therefore, leads to a complex final state characterized by reconstructed jets, missing energy from undetected neutrinos and leptons.

1.2.2 Current state of the experimental measurements related to Top Quark Physics at Tevatron and LHC

By now, most of the available Tevatron data has been analysed, providing strong evidence that the top quark is indeed the particle expected in the Standard Model. A complete review of top quark physics results from the Tevatron can be found at [31]. On the other hand, the LHC has operated remarkably well during 2010, 2011 and 2012, producing per experiment 835k top quark pairs at a center-of-mass energy of 7 TeV and 4760k top quark pairs at $\sqrt{s}$ = 8 TeV. The large amount of data collected has allowed for very precise experimental data analysis, that is in agreement with the Standard Model top quark. A review of the top quark physics at the LHC can be found in [32], [33].

The top quark best measured property is its mass followed by the top production cross section. The latest top mass and cross section measurements both at Tevatron and the LHC are shown in Tables 1.5 and 1.6. With more than 5 fb^{-1} of Tevatron data analysed, the combined

measurements of the DØ and CDF experiments on the top quark mass have relative uncertainty of only 0.5% [34]. ATLAS and CMS have also performed various measurements, with more than 20 fb^{-1} of data analysed, and they have published a first combined LHC result [35] that matches the Tevatron measurement although it is slightly less precise due to higher systematic uncertainties dominated by the jet calibration and the signal modelling.

	$\sqrt{s}$ (TeV)	Top mass (GeV/c^2)	$\mathcal{L}_{int}$ (fb^{-1})	Reference
Tevatron	1.96	173.20 ± 0.51 (stat.) ± 0.71 (syst.)	8.7	[34]
LHC	7	173.29 ± 0.23 (stat.) ± 0.92 (syst.)	4.7	[35]

Table 1.5: *Latest measurements of Top quark mass at Tevatron and LHC.*

	$\sqrt{s}$ (TeV)	$\sigma_{t\bar{t}}$ (pb)	$\mathcal{L}_{int}$ (fb^{-1})	Reference
CDF & DØ	1.96	7.68 ± 0.20 (stat.) ± 0.36 (syst.)	8.8	[36]
ATLAS & CMS	7	173.3 ± 2.3 (stat.) ± 9.8 (syst.)	0.7-1.1	[37]
CMS dilepton	8	239 ± 2 (stat.) ± 11 (syst.) ± 6 (lumi.)	5.3	[38]
ATLAS ℓ+ jets	8	241 ± 3 (stat.) ± 31 (syst.) ± 9 (lumi.)	5.8	[39]

Table 1.6: *Top pair production cross section latest measurements at Tevatron and LHC.*

Figure 1.3 shows a summary of top pair cross section measurements done by the DØ and CDF experiments at Tevatron for diverse final states [36]. The final combination measurement agrees with the theoretical expectation (yellow band) within systematic uncertainties. All cross section measurements at Tevatron and LHC are in good agreement with the evolution with the center-of-mass energy predicted by the theory and shown in Figure 1.2.

The ATLAS experiment uses all decay channels to perform a cross section measurement: events in the single lepton channel with b-tagging [40], single lepton channel without b-tagging [41], dilepton channel including only electrons and muons [42], [43], and all-hadronic channel [44]. Similar measurements have been performed by the CMS collaboration [45], [46], [47]. These measurements are summarized in Figure 1.4. The top quark pair production cross section measurements agree well with the theoretical calculations.

So far, experimental data are consistent with SM predictions. Searches for beyond the Standard Model (BSM) processes are being performed both in the production process as well as in the decays of top-antitop pairs.

Recent measurements of the $t\bar{t}$ forward-backward production asymmetry from the DØ [48] and CDF [49] experiments at the Tevatron indicate possible disagreement with SM expectations. Measurements of the $t\bar{t}$ charge asymmetry at the LHC with 5.0 fb^{-1} of 7 TeV collision data have been performed by the CMS and ATLAS detectors. CMS determined [50]:

$$A_C = -0.010 \pm 0.017\ (\text{stat.}) \pm 0.008\ (\text{syst.}), \tag{1.6}$$

while the ATLAS Collaboration measured [51]:

$$A_C = 0.006 \pm 0.010\,(\text{total}). \tag{1.7}$$

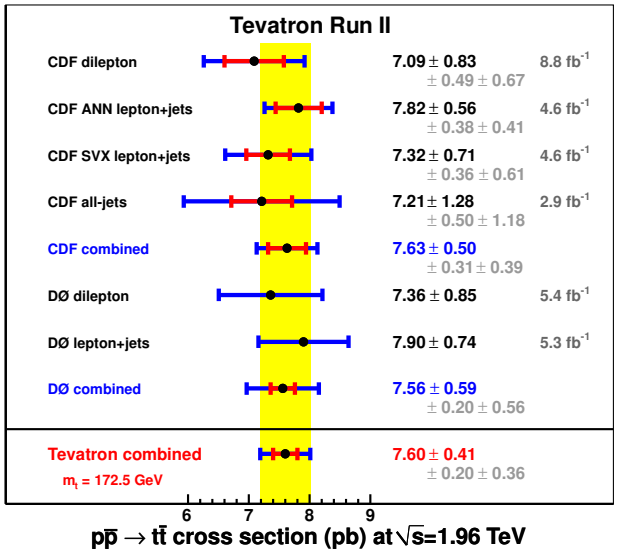


Figure 1.3: *Tevatron Run II top pair cross section measurements.*

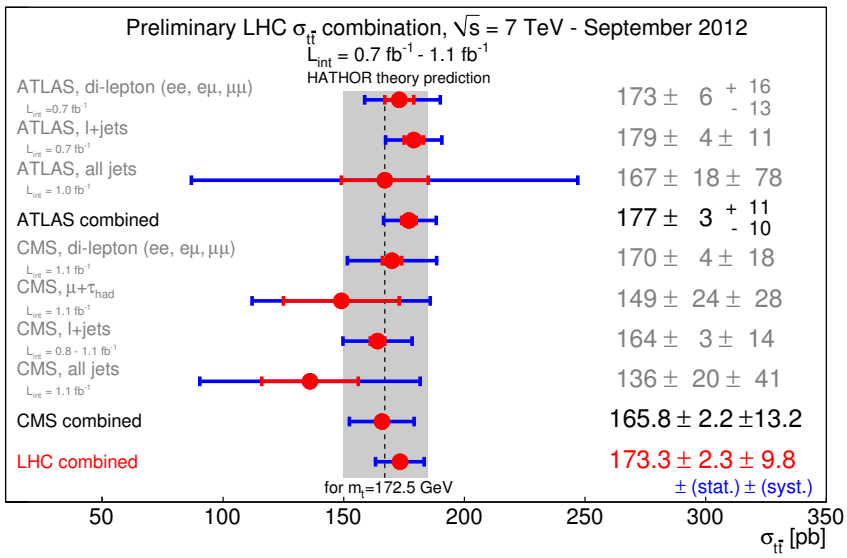


Figure 1.4: *Top pair cross section measurements by the ATLAS and CMS collaborations and the result of the LHC combination for a center-of-mass energy of* 7 TeV. *All measurements agree with the theoretical prediction (vertical grey band).*

Both measurements are consistent with the SM prediction of $A_C = 0.0123 \pm 0.0005$ [52].

SM top quark pairs are produced unpolarised and their spins are predicted to be correlated. The top-pair polarisation has been recently measured by ATLAS [53] and CMS [54] and no deviation from the SM predictions has been observed. Spin Correlation in $t\bar{t}$ events has been recently observed by ATLAS [55] using a dataset of 2.1 fb^{-1} of 7 TeV collision data in the dilepton channel. CMS performed a similar measurement in the dilepton channel, using 5.0 fb^{-1} [54]. Both results are in full agreement with the SM predictions.

Top flavour-changing neutral-current (FCNC) decays are highly suppressed by the SM, with branching ratios of the order of 10^{-15}, that could be enhanced up to 10^{-3} in many BSM models. A search for flavor-changing neutral currents has been performed for top quark pair production events, with one quark decaying through the $t \rightarrow Zq$ FCNC (q=u,c) channel, and the other through the Standard Model dominant mode $t \rightarrow Wb$. Both ATLAS, with 2.1 fb^{-1} of 7 TeV data analysed [56], and CMS, with 19 fb^{-1} of 8 TeV collision data analysed [57], have not seen an excess in the observed number of events relative to the Standard Model prediction. Thus no evidence for flavor-changing neutral currents in top quark decays has been found so far.

1.3 Importance of Tau Leptons in Top Physics

Tau leptons play an important role in electroweak measurements and studies of the top quark cross section. Within the Standard Model, measurement of the $t\bar{t}$ production cross-section in the tau plus jets and tau plus lepton channels provides a cross-check of measurements made in the electron and muon channels. This thesis will focus on the measurement of the $t\bar{t}$ production cross section in the lepton plus tau channel.

In table 1.7 all $t\bar{t}$ decay modes with their respective branching ratios are shown. The decaying channel $t\bar{t} \rightarrow bW(e/\mu + \nu_{e/\mu})bW(\tau_{had} + \nu_\tau)$ (highlighted in yellow in table 1.7) represents only a 6% of the total, which makes the cross section measurement statistically limited when compared to other channels. Nevertheless, it will be an independent measurement and it will contribute to the reduction of the uncertainty of the combined measurement of all channels.

Tau production has also been predicted as a possible signature for certain beyond the Standard Model Higgs bosons, various SUSY models and other potential new physics.

Tau leptons have a mass of $m_\tau = 1776.99^{+0.29}_{-0.26}$ MeV and, unlike electrons, that are stable, tau leptons have a lifetime of $\tau = (290.6 \pm 0.1) \times 10^{-15}$s ($c\tau = 87\mu m$), short enough to decay before reaching the detector. Their production and decays are well spaced in time and their decaying modes have been well established in low energy experiments, making them an ideal signal in new physic searches. For example, if a charged Higgs exists as predicted by the MSSM, and top quarks could decay to a charged Higgs and a b quark with the $H^\pm$ decaying to a τ lepton and a neutrino, an excess in the $\ell + \tau$ channel over the other dilepton channels would be potentially observed and the ratio

$$\mathrm{R} = \frac{t \rightarrow \tau\nu_\tau}{t \rightarrow \ell\nu_\ell} \quad (\ell = e, \mu) \tag{1.8}$$

will be greater than 1, which is the value predicted by the Standard Model. Therefore, obtaining a value of R greater than 1 would constitute a sign of new physics, providing experimental evidence of the the existence of a charged Higgs boson.

Category	Decaying Mode	branching ratio (BR)	
dileptonic	$t\bar{t} \to e\nu b e\nu \bar{b}$	1/81	4/81 (5%)
	$t\bar{t} \to \mu\nu b \mu\nu \bar{b}$	1/81	
	$t\bar{t} \to e\nu b \mu\nu \bar{b}$	2/81	
	$t\bar{t} \to e\nu b \tau\nu \bar{b}$	2/81	5/81 (6%)
	$t\bar{t} \to \mu\nu b \tau\nu \bar{b}$	2/81	
	$t\bar{t} \to \tau\nu b \tau\nu \bar{b}$	1/81	
1 Lepton + jets	$t\bar{t} \to q\bar{q} b e\nu \bar{b}$	12/81	24/81 (30%)
	$t\bar{t} \to q\bar{q} b \mu\nu \bar{b}$	12/81	
	$t\bar{t} \to q\bar{q} b \tau\nu \bar{b}$	12/81	12/81 (15%)
Full Hadronic	$t\bar{t} \to q\bar{q} b q\bar{q} \bar{b}$	36/81	36/81 (44%)

Table 1.7: *$t\bar{t}$ decaying modes and associated branching ratios.*

The much larger cross section for $t\bar{t}$ production at the LHC provides an opportunity to measure the BR($t \to \tau\nu b$) with a higher precision, and thus increase the sensitivity to $H^{\pm}$ or other processes that enhance this branching ratio.

Upper limits on the branching ratio of top quark decays to $H^{\pm}$ bosons have been published by Tevatron and LHC experiments with a 95% CL, assuming that the branching ratio of the charged Higgs boson to a τ lepton and a neutrino is 100%. Tevatron studies, with 1 fb^{-1} of D$\emptyset$ data analysed, placed upper limits of 12-26% on the $t \to H^{+}b$ branching ratio for charged Higgs boson masses between 80 and 155 GeV [58]. The analysis based on 4.6 fb^{-1} of pp collision data at $\sqrt{s}$ = 7 TeV collected by ATLAS leads to significant improvement in the branching ratio upper limits, with values between 5% and 1% for charged Higgs boson masses of 90 and 160 GeV, respectively [59]. CMS studies with 2.2 fb^{-1} of pp collision data at $\sqrt{s}$ = 7 TeV have set the upper limits on the branching ratio in the range of 2-4% for masses between 80 and 160 GeV [60]. These results constitute a significant improvement compared to the limits provided by Tevatron.

Recent analysis of 19.5 fb^{-1} of pp collision data at $\sqrt{s}$ = 8 TeV collected by ATLAS in 2012 have set the limits of the on the $t \to H^{+}b$ branching ratio in the range 0.24%-2.1% with a 95% CL for the mass range 90 < $m_{H^{+}}$ < 160 GeV. For the mass range 180 < $m_{H^{+}}$ < 600 GeV, 95% CL upper limits are set on the production cross section of a charged Higgs boson in the range 0.017-0.9 pb [61].

In the context of the m_h^{max} scenario of the MSSM [62], ATLAS results based on 4.6 fb^{-1}

of pp collision data at $\sqrt{s} = 7$ TeV collected in 2011 show that values of $\tan\beta$ above 12-26, as well as between 1 and 2-6, can be excluded for charged Higgs boson masses between 90 and 150 GeV with a 95% CL. The analysis of 19.5 fb^{-1} of pp collision data at $\sqrt{s} = 8$ TeV collected by ATLAS leads to an exclusion of $\tan\beta > 1$ for light charged Higgs masses in the range $100 < m_{H^+} < 140$ GeV with a 95% CL. For heavy charged Higgs bosons values of $\tan\beta$ between 47 and 63 have been excluded in the mass range $200 < m_{H^+} < 300$ GeV.

1.3.1 Top pair production cross section in the $\ell + \tau$ channel

The current best top pair production cross section measurements in the dilepton channel with tau lepton from the Tevatron and LHC collaborations are shown in Table 1.8. The best Tevatron measurement of the $t\bar{t} \to \ell + \tau$ cross section has an uncertainty of 25% [63], [64], [65]. While it agrees with SM predictions, the errors are notably large. To improve this result we need to better estimate the background contributions. The large data samples for $t\bar{t}$ production at the LHC provide an opportunity to measure $\sigma_{t\bar{t}}$ using final states with an electron or a muon and a τ lepton with high precision. The CMS collaboration has measured it with a 18% precision [66].

The ATLAS collaboration has published a first measurement of the $t\bar{t} \to \ell + \tau$ cross section with a 14% precision [67]. An alternative method, that aims to improve the background description and reduce statistical uncertainties is developed in Chapter 5.

	$\sqrt{s}$ (TeV)	$\sigma_{t\bar{t}\to\ell+\tau}$ (pb)	$\mathcal{L}_{\text{int}}$ (fb^{-1})	Reference
CDF	1.96	8.2 ± 2.3 (stat.) ± 1.2 (syst.) ± 0.5 (lumi.)	9.0	[65]
DØ	1.96	$7.32\ ^{+1.34}_{-1.24}$ (stat.) $^{+1.49}_{-1.31}$ (syst.) ± 0.39 (lumi.)	1.2	[63]
ATLAS	7	186 ± 13 (stat.) ± 20 (syst.) ± 7 (lumi.)	2.05	[67]
CMS	7	143 ± 14 (stat.) ± 22 (syst.) ± 3 (lumi.)	2.2	[66]

Table 1.8: *Top pair production cross section measurements at Tevatron and LHC in the lepton + tau channel.*

2

Experimental Setup: The LHC and the ATLAS detector

2.1 The LHC

The Large Hadron Collider (LHC) [68], is a superconducting proton-proton collider built by the European Organization for Nuclear Research (CERN) [69] on the French-Swiss border. It is located at a depth between 50 m and 175 m and consists of a 27 km ring of superconducting magnets with a number of accelerating structures to boost the energy of the particles along the way. A scheme of the accelerator complex is shown in Figure 2.1.

Packets, or bunches, of protons are produced from the ionization of hydrogen atoms and then sent through a series of pre-accelerators. First, the protons are accelerated up to 50 MeV in the LINAC (LINear particle ACcelerator). Then, in the Proton Synchrotron Booster (PSB), they are accelerated to 1.4 GeV. In the Proton Synchrotron (PS), the proton energy is increased up to 26 GeV. In the last step of the pre-acceleration, the protons are accelerated up to 450 GeV[1] in the Super Proton Synchrotron (SPS). At this stage, two sets of proton beams are injected to the main ring in opposite directions. Superconducting dipole and quadrupole magnets with a field of up to 8.3 T keep the protons aligned and stable in the two rings.

Once the beams are injected into the LHC main ring, their energy is increased to the desired level by ramping up the magnets from 0.54 T to 8.3 T, and once this energy is maintained the beams are declared stable. Then the experiments start the data taking. The beams circulate through the ring and collide at certain points where the detectors are situated during periods of time up to 24 hours before the beam losses stability or the loss of protons in the collisions renders the quality of the beams insufficient. At this point the beam is directed out the accelerator, "dumped" into a large metal block where it is absorbed, and the magnets are ramped down to

[1] This is the minimum energy at which the LHC can maintain a stable beam.

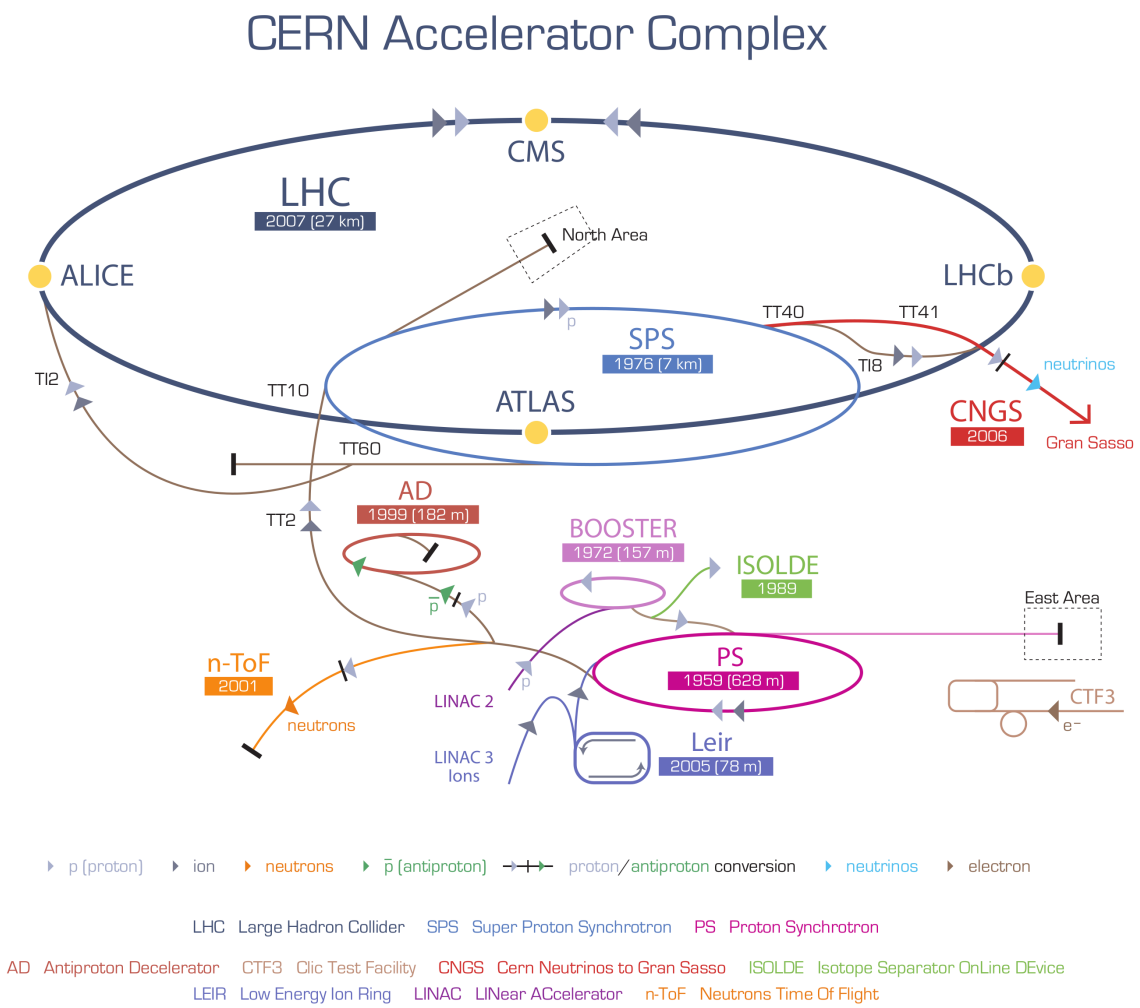


Figure 2.1: *Illustration of the CERN accelerator complex.*

0.54 T.

An important feature of the LHC operation is the luminosity, which measures the number of collisions produced by the accelerator. The instantaneous luminosity provides a measurement of the rate of collisions:

$$\frac{dN}{dt} = \mathcal{L}_{\text{ins}} \cdot \sigma_{\text{event}}, \tag{2.1}$$

where $\frac{dN}{dt}$ is the event production rate in the LHC collisions, σ_{event} is the cross section of the

event under study and $\mathcal{L}_{\text{ins}}$ is the machine instantaneous luminosity, which depends only on the beam parameters and can be written as

$$\mathcal{L}_{\text{ins}} = f \cdot \frac{N_b^2 n_b \gamma_r}{\varepsilon_n \beta^* A}, \tag{2.2}$$

where f is the frequency of bunch crossings, N_b the number of particles per bunch, n_b the number of bunches per beam, γ_r the relativistic gamma factor, ε_n the normalized transverse beam emittance, β^* the optical beta function at the collision point, and A is the effective cross-sectional area of the beams, with the angles of the beam collision taken into account.

The design values of N_b, n_b, β^*, ε_n and $\mathcal{L}_{\text{ins}}$ for the proton-proton collisions are shown in Table 2.1. The LHC is designed to collide proton beams up to 14 TeV of center of mass energy and a luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$. It can also collide heavy ions (Pb) with an energy of 2.78 TeV per nucleon and a luminosity of $10^{27}\text{cm}^{-2}\text{s}^{-1}$.

The integrated luminosity, on the other hand, measures the total number of collisions. Reaching higher instantaneous luminosities lead to a larger number of interactions per beam crossing, and consequently a larger integrated luminosity collected by the experiments in a certain period of data taking.

Achieving high luminosity collisions is of extreme importance since to explore new physics and rare phenomena, the LHC requires not only high energies but also an extremely high collision rate, since the probability of producing rare SM processes or hypothetical new particles in any given proton-proton interaction is really small.

Therefore, luminosity is one of the most important parameters that influences the performance of the LHC. There are several ways to maximize the luminosity: increase the number of bunches or the number of protons per bunch or reduce beam size, the emittance or the β^*. During its first three years of operation the LHC has managed an impressive performance, increasing the number and energy of the bunches and reducing β^*, achieving a 77% of the design luminosity.

The LHC was designed to produce $\sqrt{s} = 14$ TeV collisions. During the initial run, in 2008, an electrical failure in the interconnection of two of the super-conducting magnets lead to an explosion that damaged several other magnets. Problematic connections were found and repaired in several other magnets and prevention measurements to diagnose and hopefully avoid similar problems were established. It was also decided to run the LHC at a reduced energy until further repairs and improvements could be performed [70].

On November 2009 proton beams were successfully circulated again and the first proton-proton collisions were produced at the injection energy of 450 GeV per beam. In 2010 the LHC produced pp collisions with a center of mass energy of 7 TeV. In 2011 the LHC reached an instantaneous luminosity of $\mathcal{L}_{\text{ins}} = 3.65 \cdot 10^{33}\text{cm}^{-2}\text{s}^{-1}$ and delivered a total of 5.46 inverse femtobarns ($\mathcal{L}_{\text{int}} = 5.46\ \text{fb}^{-1}$) to the experiments. The 4th of April of 2012 the LHC 2012 run started at a beam energy of 4 TeV, corresponding to a collision energy of 8 TeV. The total data delivered by the LHC in 2012 was 22.8 fb^{-1}, four times larger than the 2011 run and higher than the luminosity Tevatron accumulated over 10 years. Table 2.1 shows the LHC running conditions for the nominal design and their evolution during the run I data taking period (2010-2012) (see references [71] and [72] for more details).

In 2013 the LHC entered a long technical stop to prepare for running at its full design energy of around 7 TeV per beam and a luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$ during Run II, which will be almost double of the energy and luminosity of Run I. At the end of Run II, the LHC will undergo another upgrade to deliver two to three times the instantaneous luminosity at $\sqrt{s} = 14$ TeV during Run

Parameter	Nominal	2010	2011	2012
Beam Energy [TeV]	7	3.5	3.5	4
Bunch Spacing [ns]	25	150	50	50
Number of bunches	2808	368	1380	1380
Number of particles per bunch	$1.15 \cdot 10^{11}$	$1.2 \cdot 10^{11}$	$1.45 \cdot 10^{11}$	$1.7 \cdot 10^{11}$
β^* [m]	0.55	3.5	1.5/1	0.6
ε_n [μm rad]	3.75	2.4	2.4	2.5
Relativistic γ	7461		3730	
Luminosity [$\mathrm{cm^{-2}s^{-1}}$]	10^{34}	$2.1 \cdot 10^{32}$	$3.7 \cdot 10^{33}$	$7.7 \cdot 10^{33}$
Total Luminosity delivered [$\mathrm{fb^{-1}}$]	-	0.048	5.46	22.8

Table 2.1: *LHC running conditions for the nominal design and during the run I data taking. Each parameter value corresponds to the best performance achieved during the year.*

III. The High-Luminosity (HL-LHC) upgrade will take the LHC to its full physics potential. The targeted luminosity of $\sim 5 \cdot 10^{34}\mathrm{cm^{-2}s^{-1}}$ is well beyond its original design value. In order to properly operate at such luminosity values, the current experiments will require numerous upgrades and replacements.

Table 2.2 shows a summary of the LHC upgrade plans for the following decade. A more detailed planning of the upgrades of the LHC machine can be found in [73]. Both the increased energy and the increased luminosity will significantly extend the reach and therefore the opportunities to explore new physics.

	Beam Energy [TeV]	Luminosity [$\mathrm{cm^{-2}s^{-1}}$]	Luminosity [$\mathrm{fb^{-1}}$]
Run 2 (2015-2017)	6.5-7	$\sim 1 \cdot 10^{34}$	100
Run 3 (2018-2021)	7	$\sim 2 \cdot 10^{34}$	350
HL-LHC (2023-?)	7	$\sim 5 \cdot 10^{34}$	3000

Table 2.2: *LHC upgrade plans to design luminosity and beyond.*

2.1.1 The LHC experiments

There are four main experiments in the LHC (ATLAS, CMS, LHCb and ALICE), installed in the interaction points shown in Figure 2.1, and two smaller experiments (TOTEM and LHCf) located along the LHC ring.

ATLAS [74] [75] (A Large Lhc Toroidal ApparatuS) and **CMS** [76] (Compact Muon Selenoid) are both multipurpose experiments design to test the Standard Model and search for signs of new physics. They are high luminosity experiments, both aiming at a peak luminosity of $10^{34}\mathrm{cm^{-2}s^{-1}}$ for proton operation. Having two independently designed detectors is vital for cross-confirmation of any new discoveries made. The data analysed for this thesis has been collected with the ATLAS detector. A more detailed description of the detector components and its relevance for the analysis can be found in section 2.2.

The **LHCb** detector [77] is an experiment dedicated to heavy flavour physics at the LHC. Its

Detector Component	Resolution	E/p_{T} range	η coverage
Inner Detector Tracker	$\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} = 0.05\% p_{\mathrm{T}} \oplus 1\%$	$p_{\mathrm{T}} < 1$ TeV	±2.5
EM Calorimeter	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$	$10 < E < 245$ GeV	±3.2
Hadronic Calorimeter:			
barrel and end-cap	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$	$10 < E < 300$ GeV	±3.2
forward	$\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$		$3.1 < \lvert\eta\rvert < 4.9$
Muon Spectrometer	$\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} = 10\%$	$p_{\mathrm{T}} = 1$ TeV	±2.7
	$\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} = 4\%$	$p_{\mathrm{T}} < 200$ GeV	±2.7

Table 2.3: *Designed performance goals of the ATLAS sub-detectors. For high-p_{T} muons, the muon spectrometer performance is independent of the inner detector system. The units for E and p_{T} are in* GeV.

primary goal is to look for indirect evidence of new physics in CP violation and rare decays of beauty and charm hadrons. As of summer of 2013, no significant deviation from the Standard Model prediction has been found.

ALICE (A Large Ion Collider Experiment) is a dedicated heavy-ion detector designed to exploit the unique physics potential of nucleus-nucleus interactions at LHC energies. The aim of ALICE [78] is to study the physics of strongly interacting matter at extreme values of energy density and temperature, where the formation of a new phase of matter, the quark-gluon plasma (which is believed to have existed soon after the Big Bang), is expected. For this purpose, the LHC has provided collisions between lead ions. In August 2012 ALICE scientists announced that their experiments produced quark-gluon plasma with temperature at around 5.5 trillion degrees, the highest temperature mass achieved in any physical experiments thus far.

The **TOTEM** (TOTal Elastic and diffractive cross section Measurement) experiment [79] is dedicated to measure the effective size of the proton and also monitor accurately the LHC's luminosity. It measures the total pp cross-section with a luminosity-independent method and study elastic and diffractive scattering at the LHC.

The **LHCf** (LHC forward detector) [80] use very forward neutral particles emitted in LHC collisions as a source to simulate cosmic rays in laboratory conditions. Analysing the cascades of charged particles within the LHC in analogy to the cosmic rays that bombard the Earth from outer space, helps to interpret and calibrate studies of the cosmic rays. Research on the highest energy cosmic-rays (> 1019 eV) has great scientific interest since their origin, propagation and interactions are unknown and may yield information about new physics.

2.2 The ATLAS detector

ATLAS is as a general-purpose detector. Its design, assembly, calibration, and overall preparation have been a collaborative effort among the 37 nations and 2500 scientists who contribute to the ATLAS project.

The layout structure, with the inner detector, calorimeters, muon spectrometer and magnet system is shown in Figure 2.2. The designed energy and momentum resolutions of the ATLAS sub-detectors [81] are summarized in Table 2.3.

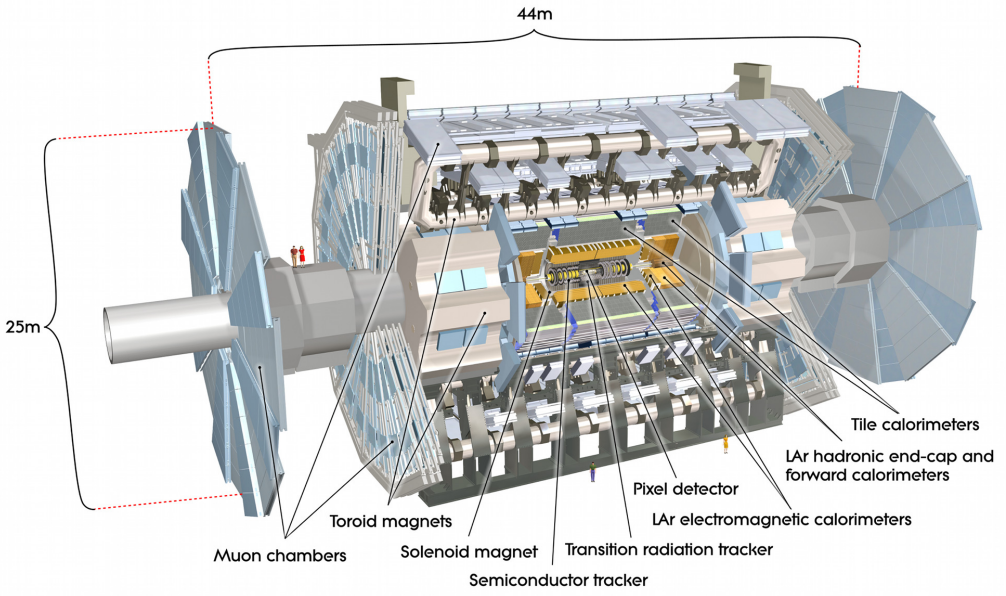


Figure 2.2: *The ATLAS detector layout with all its sub-detectors labelled.*

2.2.1 ATLAS layout

The ATLAS detector is 22 m high, 44 m long and weights 7000 tons. The detector is built symmetrically around the LHC beam line with the particle interaction point in its center. It consists of four major components arranged in layers forming a cylindrical structure to allow the maximum spatial coverage of the particles produced in each collision. Radially, from the inside to the outside, it consists of an inner tracking detector that measures the momentum of charged particles embedded into a solenoid magnet, a system of electromagnetic and hadronic calorimeters to measure the energies carried out by the particles and, finally, a muon spectrometer which identifies and measures the momentum of muons situated inside a large toroidal magnet.

The only stable particles that escape detection are the neutrinos. Their presence is inferred by noticing a momentum imbalance among detected particles. To achieve this, the detector must be "hermetic", and detect all non-neutrinos particles produced, with no blind spots. This is one of the greatest engineering challenges of the ATLAS detector.

Spacial coordinates in this thesis are given using the right-handed ATLAS coordinate system (η,ϕ,z). The nominal interaction point is defined as the origin of the coordinate system. The counterclockwise beam direction defines the z-axis, and the x-y plane is transverse to the beam direction. The positive x-axis is defined as pointing from the interaction point to the center of the LHC ring and the positive y-axis as pointing upwards. The pseudorapidity η is defined as $\eta = -\ln(\tan(\theta/2))$, where the polar angle θ is measured with respect to the LHC beamline (see Figure 2.3). The azimutal angle ϕ is measured with respect to the x-axis. Angular separation between points is defined by

$$\Delta R = \sqrt{(\eta' - \eta)^2 + (\phi' - \phi)^2} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \quad (2.3)$$

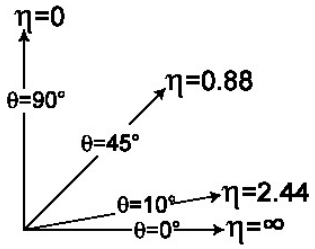


Figure 2.3: *Pseudorapidity (η) values shown on a polar plot. In particle physics, an angle of zero is usually along the beam axis, and thus particles with high pseudorapidity values are generally lost, escaping through the space in the detector along with the beam.*

ATLAS was built to be as symmetric in the ϕ coordinate as possible. Therefore, only the η coordinate is usually used to label the regions covered by the detector. Large values of $|\eta|$ are close to the beam line, and are often referred to as the forward part of the detector, and, similarly, smaller values of $|\eta|$ are perpendicular to the beam and are often referred to as central. The part of the detector corresponding to $z > 0$ is called "Side A", while for $z < 0$ is called "Side C". Some subsystems of the detector are composed with detecting elements parallel to the z-axis, forming the so-called barrel of the detector. On the two sides of the barrel, detecting elements are arranged in transversal planes to the beam axis, forming the end-caps.

Inner Detector

The Inner Detector (ID) begins a few centimetres from the proton beam axis, extends to a radius of 1.05 m, and it is 7 m long in the direction of the beam pipe. Its basic function is to track charged particles in the range $|\eta| < 2.5$ by detecting their interaction with material at discrete points, revealing detailed information about the type of particle and its momentum. The magnetic field surrounding the entire inner detector causes charged particles to curve; the direction of the curve reveals a particle's charge and the curvature radius reveals its momentum. Another role of the inner detector is to reconstruct primary vertices for the identification of the collision vertex and particles such as electrons or jets from b-quark hadronization.

The Inner Detector consists of three independent but complementary sub-detectors: two precision tracking detectors, the pixel and the semiconductor tracker (SCT), and the straw tubes of the transition radiation tracker (TRT):

- The ATLAS Pixel Detector is the the innermost part of the detector. It provides a very high granularity and high precision set of measurements as close to the interaction point as possible. The Pixel Detector contains three layers in the barrel and three disks on each end-cap, with a total of 1,744 modules, each measuring 2 by 6 cm. The detecting material is 250 μm thick silicon. Each module contains 16 readout chips and other electronic components. The smallest unit that can be read out is a pixel, each one of them 50 by 400 μm^2; there are roughly 47,000 pixels per module. In total, the Pixel Detector has

over 80 million readout channels, which is about 50% of the total readout channels. The intrinsic resolution is 10 μm in the R$-\phi$ direction and 115 μm in the z axis (barrel) and R (end-caps).

- The SCT is the middle component of the inner detector. It is similar in concept and function to the Pixel Detector but with long, narrow strips rather than small pixels, making coverage of a larger area practical. Each strip measures 80 μm by 12.6 cm. The SCT is the most critical part of the inner detector for basic tracking in the plane perpendicular to the beam, since it measures particles over a much larger volume than the Pixel Detector, with more sampled points and roughly equal (albeit one dimensional) accuracy. It is composed of four double layers of silicon strips in the barrel and nine disk-shape layers in the endcap, and has 6.2 million readout channels and a total area of 61 m^2. It is aligned so that the charged particles can cross more than 8 layers in total.

- The TRT detector is formed by layers of gaseous straw tube elements interleaved with transition radiation material. The drift tubes, of 4 mm diameter, are especially developed to have good electrical and mechanical properties with minimal wall thickness. Straw tubes with a length of 144 cm are arranged in the barrel region parallel to the z-axis forming up to 73 layers of straws interleaved with polypropylene fibres. In the end-cap region, straws with a length of 37 cm are placed radially in wheel segments interlaced with foil. The straw tubes are superimposed so that any charged track with $p_{\mathrm{T}} > 0.5$ GeV can cross at least 36 layers. The straw tubes are filled with a gas mixture of Xe, CO_2 and O_2 and are read out with an anode of gold plated tungsten. These dispositions provide transition radiation for electron identification. The TRT has a total of 351,000 readout channels. It only provides R$-\phi$ information with an intrinsic accuracy of 130 μm per tube, but contributes significantly to the momentum measurement due to the higher number of track points and longer track length than the silicon detectors.

The general function of each system is to record a "hit" when a particle passes through a certain point of the detector. These hits, when fit together, form the particle's trajectory. The intrinsic resolution of the three sub-detectors is shown in Table 2.4.

Pixel Detector ($R\phi/z$)	SCT ($R\phi/z$)	TRT ($R\phi$)	Ref.
10 μm / 115 μm	17 μm / 580 μm	130 μm	[81]

Table 2.4: *Intrinsic resolution of three ATLAS ID sub-detectors.*

Calorimeters

The energy of particles in ATLAS is measured by a system of calorimeters in the barrel and end-cap regions. The calorimeters are situated outside the solenoidal magnet that surrounds the inner detector, with full ϕ-symmetry and coverage around the beam axis. They cover a pseudorapidity range of $|\eta| < 4.9$. Their purpose is to measure the energy of charged and neutral particles by absorbing it. The calorimeters also provide input to the ATLAS trigger system described in Section 2.3.1.

ATLAS sampling calorimeters use alternating layers of absorber and active material. The incident particle interacts with the material in the calorimeter producing a shower of particles that are detected in the active medium which measures the energy deposited on the material. In general, the absorber medium is chosen to be a dense material with a high stopping power so that particles in the energy range under study are contained in the calorimeter. Liquid argon (LAr) was chosen as the active material for most of the calorimeters due to its resistance to radiation, its linear behaviour in a wide energy range and its stability of response. Some parts of the calorimeter also use plastic scintillators.

This way, the calorimeter system is designed to stop all the interacting particles emerging from the interaction point, except muons, measuring their energy and position.

There are two basic calorimeter systems: an inner electromagnetic (EM) calorimeter and an outer hadronic calorimeter. The barrel region of the EM calorimeter provides fine granularity for precision measurements of electrons, positrons and photons. The barrel and end-cap of the EM calorimeter, along with the hadronic calorimeters provide appropriate jet and missing transverse energy measurements.

- The LAr Electromagnetic Calorimeter is situated outside of the inner detector. It is built using a sampling technique, utilizing LAr as the active detector medium and lead plates as the passive absorber material. The lead plates are arranged in layers with the LAr sandwiched in between, both in the barrel ($|\eta| < 1.475$) and the two end-cap components ($1.375 < |\eta| < 3.2$). It is built in an accordion geometry providing a complete ϕ symmetry without azimuthal cracks. The electromagnetic calorimeter absorbs energy from particles that interact electromagnetically, which include charged particles and photons. The showers in the argon liberate electrons that are later collected and recorded. It has high precision, both in the amount of energy absorbed and in the precise location of the energy deposited. The EM calorimeter has a fine granularity, with a typical value of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$.

- The Hadronic Calorimeter absorbs energy from particles that pass through the EM calorimeter, but do interact via the strong force; these particles are primarily hadrons. It is less precise, both in energy magnitude (see Table 2.3) and in the localization, with a typical granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$. It is composed of three sub-detectors: a Tile calorimeter, a liquid-argon hadronic end-cap calorimeter (HEC) and the liquid-argon forward calorimeter (FCal):

  - The Tile Calorimeter is placed directly outside the EM calorimeter and cover a range of $|\eta| < 1.7$. It is a sampling calorimeter with plastic scintillating tiles as the active material and steel as absorber. Hadronic showers developing in the calorimeter illuminate the tiles and the light produced is detected and recorded.
  - The HEC consist of four wheels, two per end-cap, built from parallel copper plates with 8.5 mm LAr gaps, providing the active medium for the sampling calorimeter. The HEC is situated directly behind the end-cap electromagnetic calorimeter sharing the same LAr cryostat. It overlaps with the forward calorimeter in the range of $3.1 < |\eta| < 3.2$ and also slightly with the Tile calorimeter for $1.5 < |\eta| < 1.7$.
  - The FCal calorimeter is situated in the calorimeter region near the beam pipe ($3.1 < |\eta| < 4.9$). It consists of three modules in each end-cap, a copper one designed

for electromagnetic measurements and two tungsten ones to measure predominantly the energy of hadronic interactions. In each module there are gaps filled with LAr, which is used as the active medium.

The Muon System

The Muon Spectrometer (MS) forms the outer part of the ATLAS detector and is designed to detect charged particles exiting the barrel and end-cap calorimeters and to measure their momentum in the pseudorapidity range $|\eta| < 2.7$. It is also designed to trigger on these particles in the region $|\eta| < 2.4$. Muons are the only detectable particles that can traverse all the calorimeter absorbers without being stopped.

The Muon Spectrometer surrounds the calorimeter and measures muon trajectories to determine their momenta and charge with high precision. This measurement occurs inside a volume of magnetic field produced by superconducting toroid magnets. The detection elements, i.e., the muon chambers, are made of thousands of metal tubes equipped with a central wire and filled with gas. When the muon passes through these tubes, it leaves a trail of electrically charged ions and electrons which drift to the sides and center of the tube. By measuring the time it takes for these charges to drift from the starting point, it is possible to determine the position of the muon as it passes through.

A momentum resolution below 4% is achieved for muons with $p_{\mathrm{T}} < 200$ GeV, increasing to 10% for 1 TeV muons. The timing resolution is of the order of nanoseconds.

The MS is equipped with four different kind of chambers: Monitored Drift Tube chambers (MDT), Cathode Strip Chambers (CSC), Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC). The first two are used for precision measurements of muon tracks, while the other two are used for the trigger and data acquisition system.

The Forward Detectors

In addition to the main ATLAS detector systems described before, three smaller sets of detectors are built to provide good coverage in the very forward region:

- The first system is a Cerenkov detector called LUCID (LUminosity measurement using Cerenkov Integrating Detector), which is primarily dedicated to online luminosity monitoring. Its main purpose is to detect inelastic pp scattering in the forward direction, in order to both measure the integrated luminosity and to provide online monitoring of the instantaneous luminosity and beam conditions.

- The second system is the Zero-Degree Calorimeter (ZDC). Its main purpose is to detect forward neutrons in heavy-ion collisions.

- The most remote detector is the absolute luminosity detector ALFA (Absolute Luminosity For ATLAS). The absolute luminosity is determined via elastic scattering at small angles, using the optical theorem to connect the elastic-scattering amplitude in the forward direction to the total cross-section to extract luminosity.

The Magnet System

The magnet system of the ATLAS detector provides the magnetic field for bending the trajectories of charged particles for momentum measurement, which is determined by the radius of curvature of the tracks left within the detector.

The magnetic field inside the ATLAS detector is provided by a solenoid and three toroidal systems. The solenoid provides a homogeneous 2 T magnetic field for the inner detector. Since the solenoid magnet is situated between the inner detector and the calorimeter, it was designed to keep the material thickness as low as possible so its presence has a small impact on the energy measurement in the calorimeters.

Three air-core toroidal systems provide the bending field for the muon spectrometer. This magnetic field is highly non-uniform and its values vary between 0.2 T and 2.5 T for the barrel and up to 3.5 T for the end-cap magnets.

2.2.2 ATLAS Performance during LHC run I

Since the start of data taking in 2009, both LHC and ATLAS have been performing well beyond expectation. Before the ongoing technical shut down that started in 2013, the LHC delivered 5.5 fb^{-1} of pp collision data with a center-of-mass energy of $\sqrt{s}$ = 7 TeV in 2010 and 2011 and 22.8 fb^{-1} with a center-of-mass energy of $\sqrt{s}$ = 8 TeV in 2012. The ATLAS detector recorded a total of 26 fb^{-1}, which corresponds to a data taking efficiency higher than 90% [82].

The total integrated luminosity and data taking efficiency during stable beams and pp collisions at the LHC run I are shown in Figure 2.4. The cumulative luminosity versus time delivered to ATLAS is shown in green. It accounts for the luminosity delivered from the start of stable beams until the LHC requests ATLAS to put the detector in a safe standby mode to allow a beam dump or beam studies. The luminosity recorded by ATLAS is shown in yellow. The difference with the delivered luminosity reflects the inefficiency of the Data Acquisition System (DAQ, see section 2.3.2), as well as the inefficiency of the so-called "warm start": when the stable beam flag is raised by the LHC, the ATLAS detectors undergo a ramp of the high-voltage, ie they change their status from standby to ready, before they start taking data. The certified good quality data, which is used for physics analysis, is shown in blue and it correspond to a 90% of the total data recorded by ATLAS in 2011 and 2012. The data quality criteria requires all reconstructed physics objects to be of good data quality.

The different ATLAS subdetectors perform daily calibration runs to verify that all components are working and detect failures. These test runs are performed between fills, in the absence of beams in the LHC ring and, therefore, are not considered in the ATLAS DAQ efficiency.

The cumulative integrated luminosity recorded by ATLAS has increased every year from 45 pb^{-1} in 2010 to 5.1 fb^{-1} in 2011 and 21.7 fb^{-1} in 2012. The instantaneous luminosity has also continuously increased, reaching a maximum peak of $\mathcal{L}_{\text{ins}} = 7.73 \cdot 10^{33} \text{cm}^{-2}\text{s}^{-1}$ in 2012.

Luminosity measurement in ATLAS

The integrated luminosity is measured with ATLAS detector as the integral of the instantaneous luminosity over the given time. The instantaneous luminosity can be written as:

$$\mathcal{L} = f \cdot \frac{N_1 \cdot N_2}{A} \tag{2.4}$$

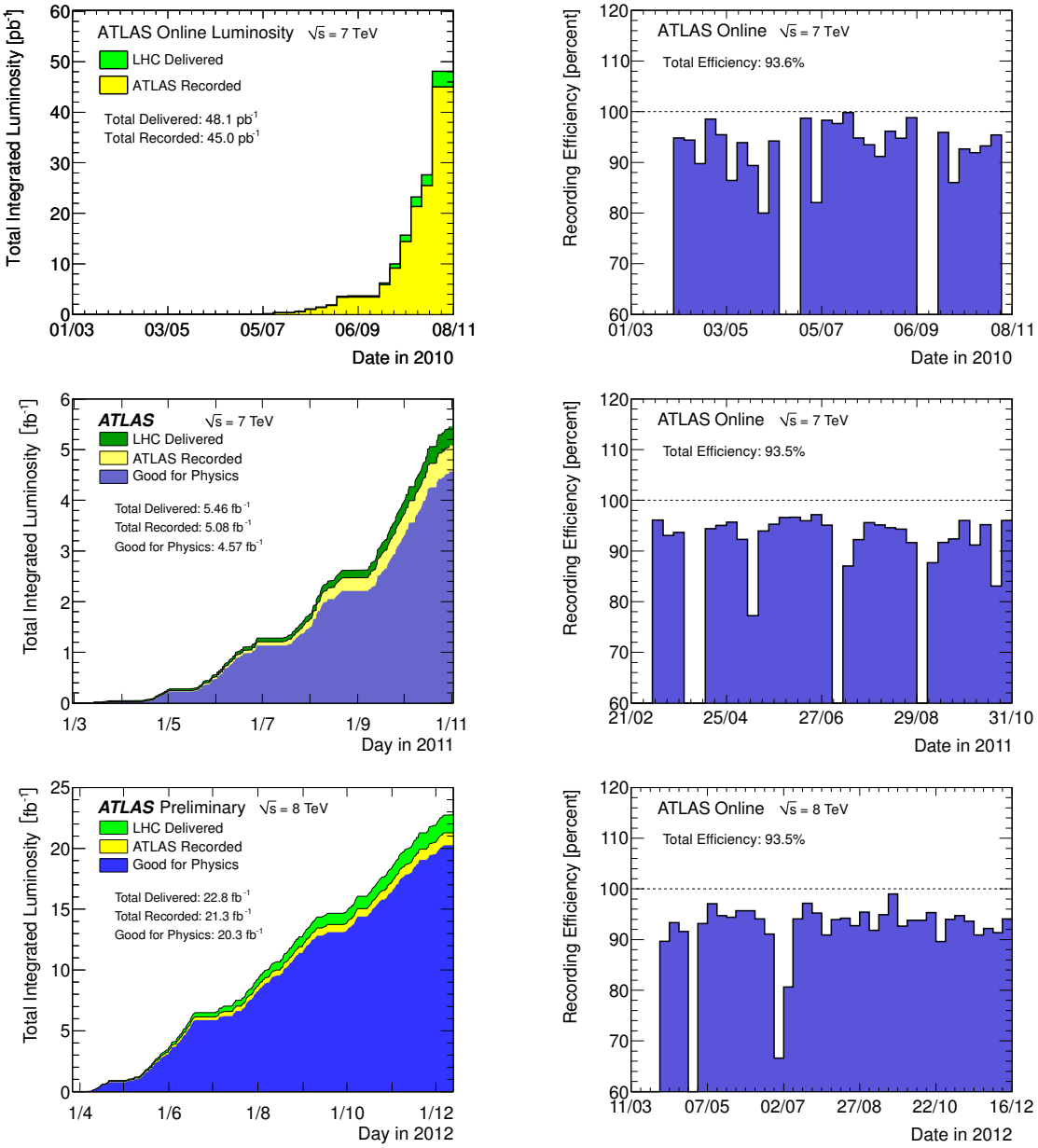


Figure 2.4: *Data taken by ATLAS detector in 2010 (top row), 2011 (middle row) and 2012 (bottom row). Left: total integrated luminosity and data quality. Right: data taking efficiency per week.*

where f is the frequency of bunch crossings, N_i is the number of protons in bunch i, and A is the effective cross-sectional area of the beams. The key aspect is to measure the transverse size of the colliding bunches (A). To do this, ATLAS uses a technique called "Van der Meer scans" to measure the beam-overlap area as well as the number of protons in each bunch.

The accuracy of the luminosity measurements in 2010 and 2011 was determined to be:

$$\frac{\Delta\mathcal{L}}{\mathcal{L}} = \begin{cases} \pm 3.5\% \text{ (2010)} \\ \pm 1.8\% \text{ (2011)} \end{cases} \tag{2.5}$$

The dominant uncertainty in the 2010 luminosity measurement comes from the bunch charge product ($N_1 \cdot N_2$), which accounts for a 3.0% of the 3.5% total uncertainty. In 2011 the main uncertainty comes from the VdM test uncertainties, with a 1.4% of a total of 1.8% [83–85].

Pile Up effect on Data

When two bunches of protons collides there is the possibility of having multiple pp interactions measured as only one event. This phenomenon is known as pileup.

There are two different sources of pileup:

1. High luminosities imply a high number of protons per bunch. Thus, the probability of multiple pp interactions in one bunch crossing increases with luminosity. This is called in-time pileup.

2. If the spacing between the bunches is shorter than the response time of the detectors, pp interactions in a previous bunch crossing can also affect the measurement. This is called out-of-time pileup. As luminosity increases at the LHC and the bunch spacing decreases, this source of pileup becomes more important.

The mean number of interactions per crossing is calculated from the instantaneous luminosity as

$$\langle\mu\rangle = \frac{\mathcal{L}_{\text{ins}} \cdot \sigma_{inel}}{n_b \cdot f_r}, \tag{2.6}$$

where $\mathcal{L}_{\text{ins}}$ is the average instantaneous luminosity over a large time period ($\Delta t \gg 600$ ns), σ_{inel} is the total cross-section of the inelastic scatterings (71.5 mb for 7 TeV collisions and 73.0 mb for 8 TeV collisions), n_b is the number of proton bunches and f_r is the LHC beam circulating frequency (11.2 kHz).

In 2010 an instantaneous peak luminosity of $\mathcal{L}_{\text{ins}} = 2.1 \cdot 10^{32}\text{cm}^{-2}\text{s}^{-1}$ was reached with a bunch spacing of 150 ns, producing an average number of 2.2 interactions per event, mostly coming from in-time pileup. In 2011, a peak instantaneous luminosity of $\mathcal{L}_{\text{ins}} = 3.65{\cdot}10^{33}\text{cm}^{-2}\text{s}^{-1}$ was achieved, and the bunch spacing was reduced to 50 ns, increasing the out-of-time pileup significantly. Then, 9 interactions per bunch crossing were observed on average. In 2012 the bunch spacing was maintained at 50 ns, and the luminosity reached a peak of $\mathcal{L}_{\text{ins}} = 7.7 \cdot 10^{33}\text{cm}^{-2}\text{s}^{-1}$, thus increasing the average number of interactions per crossing to 21. Figure 2.5 shows the luminosity weighted mean number of events per beam crossing for 2011 and 2012 pp collisions.

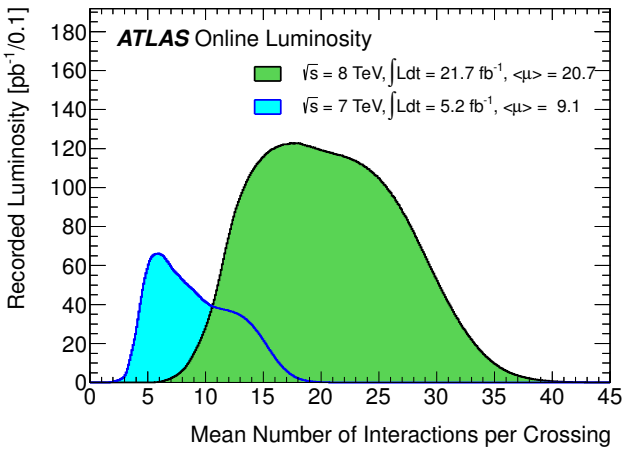


Figure 2.5: *Luminosity-weighted distribution of the mean number of interactions per crossing for the 2011 and 2012 data.*

2.3 Analysis of ATLAS data

Analysing the data taken with the ATLAS detector is a complex task. Particles are reconstructed from its decaying products, these are reconstructed from detector signatures, etc. Electronic circuits record the passage of each particle through a detector as a series of electronic signals, and send the data to the CERN Data Centre (DC) for digital reconstruction. The digitized summary is recorded as a "collision event". Physicists must sift through roughly 15 petabytes of data produced annually to determine if the collisions have thrown up any interesting physics. To digest this data we use:

1. A Trigger system to achieve the acceptable event rate and at the same time ensure that all possible signals of new physics are selected with high efficiency. It selects about 100 interesting events per second out of 1000 million others.

2. The data acquisition system (DAQ) for channelling the data from the detectors to the storage.

3. The computing system to analyse and store the data recorded.

2.3.1 ATLAS trigger system

The LHC provides proton-proton and heavy-ion collisions at high interaction rates. For protons the design beam crossing interval is 25 ns (see Table 2.1), corresponding to a crossing frequency of 40 MHz. Depending on luminosity, several collisions occur at each crossing of the proton bunches (approximately 20 simultaneous pp collisions at the nominal design luminosity of $10^{34}\text{cm}^{-1}\text{s}^{-1}$). Since it is impossible to store and process the large amount of data associated

with the resulting high number of events, a drastic rate reduction has to be achieved. This task is performed by the trigger system, which is the start of the physics event selection process.

A three-level trigger is used in ATLAS to select interesting events produced in pp collisions. Level-1 trigger is hardware-based and uses a subset of detector information to reduce the event rate to at most 75 kHz. This trigger is followed by two software-based trigger levels, level-2 and the event filter, used together to reduce the event rate to about 200 Hz.

- **Level 1 trigger** searches for high transverse-momentum muons, electrons, photons, jets, and tau leptons decaying into hadrons, as well as large missing and total transverse energy. In each event, the L1 trigger also defines one or more Regions-of-Interest (RoI's), given in coordinates η and ϕ , which are regions within the detector where the trigger selection process has identified interesting features. The RoI data include information on the type of feature identified and the criteria passed.

- **Level 2 trigger** is seeded by the RoI information provided by the L1 trigger. L2 selections use all the available detector data within the RoIs (approximately 2% of the total event data) to reduce the trigger rate to approximately 3.5 kHz, with an event processing time of about 40 ms, averaged over all events.

- The **Event filter** carries out the final stage of the event selection, reducing the event rate to roughly 200 Hz, which can be recorded for subsequent offline analysis. Its selections are implemented using offline analysis procedures within an average event processing time of the order of four seconds.

2.3.2 The ATLAS Data Acquisition System and Detector Control System

The purpose of the Data Acquisition (DAQ) system is the transport of the data belonging to a given bunch crossing, and identified by the trigger, from the detector front-end electronics to permanent storage.

After an event is accepted by the L1 trigger, the data is transferred off the detector to the ROD's (Readout Drivers). Digitised signals are formatted as raw data prior to being transferred to the DAQ system. The first stage of the DAQ, the readout system, receives and temporarily stores the data in local buffers. Event data associated with RoI's is subsequently requested by the L2 trigger. The events selected by the L2 trigger are then transferred to the event-building system and subsequently to the event filter for final selection. Events selected by the event filter are moved to permanent storage at the CERN computer centre.

In addition to the movement of data, the data acquisition also provides for the configuration, control and monitoring of the hardware and software components which together provide the data-taking functionality.

A Detector Control System (DCS) permits the coherent and safe operation of the ATLAS detector hardware. It controls, monitors and archives the operational parameters, signals any abnormal behaviour, and allows automatic or manual corrective actions to be taken. The DCS also enables bi-directional communication with the data acquisition system in order to synchronise the state of the detector with data-taking.

The DCS also handles the communication between the sub-detectors and other systems which are controlled independently, such as the LHC accelerator, the CERN technical services, the ATLAS magnets, and the detector safety system. The exchange of information between LHC

and ATLAS ensure the safe and optimal operation of the machine. The LHC machine gives ATLAS beam information. ATLAS provides information on luminosity, quality of collisions and whether or not to safely inject or dump the beams, or to move from one mode of operation to the next.

2.3.3 The computing system

The ATLAS offline computing system must support the storage, transfer and manipulation of the recorded data for the lifetime of the experiment. The system accepts real-time detector information from the data acquisition system at the experimental site; performs pattern recognition, event filtering, and data reduction and supports the physics analysis activities of the collaboration. The system also supports production and distribution of simulated data, and access to conditions and calibration information and other non-event data.

The users of the system, and the physical computer centres it comprises, are distributed worldwide, interconnected by high-speed international networks, constituting a fully distributed computing model. The system is based upon Grid technologies [86], with the common Grid services at centres defined and managed through the Worldwide LHC Computing Grid (WLCG) project [87], [88], a collaboration between LHC experiments, computing centres, and middleware providers. This way, individual institutes and national organizations finance and maintain local computing resources and a middleware software organize the computing and storage resources and the network as a virtual big computer.

The ATLAS computing model is based on a distributed tiered model [89]. CERN is the central production centre (Tier-0), responsible for processing the data coming from the trigger filters and distributing it in quasi-real time to the Tier-1 centres. CERN and the Tier-1 centres will be responsible for managing the permanent data storage and providing computational capacity for reprocessing and for analysis processes that require the access to large amounts of data. At present 10 Tier-1 centres are defined in ATLAS. Each Tier-1 centre provide services for a cloud of associated Tier-2 centres, which provide computational and storage capacity for Monte Carlo event simulation and for end-user analysis.

This computing model have been thoroughly tested and improved in the last years [90] and it has been used for the analysis presented in this thesis.

The processing and analysis of real and simulated event data is staged in several phases:

1. The first step is to collect data, triggering on events of interest. This is the raw (unprocessed) data (RDO). Despite the trigger data reduction, raw data are still huge and is impractical to repeatedly replicate or analyse it. The raw data are transferred to the CERN Tier 0 centre for further processing (in a series of data reduction steps that produce successively smaller datasets) and archiving.

2. The raw data, whether real or simulated, are reconstructed to form physical quantities such as the calorimeter clusters needed to provide the energy of electromagnetic and hadronic showers, and tracker hits to be associated to tracks whose position and momentum are to be determined. Information about particle identification (electron, photon, π^0, charged hadrons, muon) is also reconstructed from the appropriate sub-systems. Several algorithms are employed to interpret energy deposits and charged particle trajectories and exploit the different behaviour of each particle as it passes through the combined ATLAS detector. The event reconstruction results in the generation of new data: ESD data (Event

Summary Data) that is stored in the Tier-1 centres. The size of ESD data is around 2.5 MB/event.

3. Filtering ESD data produces AOD (Analysis Object Data), a compact analysis format (∼350 kB/event), designed to allow a wide range of physics analyses whilst occupying sufficiently small storage so that very large event samples may be held at many centres. AOD events contain the parameters of high-level physics objects, plus sufficient additional information to allow kinematic refitting.

4. The next step of data reduction is performed within the ATLAS software analysis framework, **Athena** [91], which uses Python as an object-oriented scripting and interpreter language to configure and load C++ algorithms and objects. The new samples produced are called D3PD (Tertiary Derived Physics Data). Contrary to the previous step, only a small fraction of the information stored in the AOD containers was recorded in the D3PD files (∼10 kB/event) and different series of datasets were produced for different analysis with different levels of object selections and event filtering and with various variable contents. The content of these D3PDs are defined by the Physics groups [90]. AOD and D3PD datasets are stored on all grid sites around the world. D3PD input datasets can be analysed using the ROOT framework [92], but their size is still too big for local storage and flat ntuples are usually created from the D3PD to perform the last steps of the analysis locally.

The analysis presented in this thesis have been done using real and simulated data samples in D3PD format customized for Top physics analyses in 2010 and 2011.

3

Samples and object selection for $t\bar{t} \to \ell + \tau$ analysis.

In this chapter we present the MC and data samples together with the particle identification that will be used in the analysis described in Chapters 4 and 5.

The signature of $t\bar{t} \to \ell + \tau$ events is characterized by the presence of two b-jets from the top quark decays, an isolated lepton (electron or muon) coming from a $W \to \ell\nu$ decay, an hadronically decaying τ lepton from the other W boson decay ($W \to \tau_{\mathrm{h}}\nu$) and missing transverse energy from the neutrinos. Any measurement performed in this final state requires a good particle reconstruction and identification from the energy deposits in the calorimeter and the charged particle trajectories in the inner detector and muon chambers. The presence of neutrinos is inferred through conservation of momentum.

The datasets and MC samples used for the analysis are specified in Section 3.1. In section 3.2 the identification of electrons, muons, jets and missing transverse energy is discussed as well as the reconstruction of tau from their decaying products signature in the detector.

The identification of tau leptons is a key aspect of the analysis presented in this thesis. Two different identification methods are detailed in Section 3.3.

3.1 Data and MC samples used in the analysis

The analysis presented in this thesis uses pp collision data with a center-of-mass energy of $\sqrt{s} = 7$ TeV acquired with the ATLAS detector. The data collected between March 30th and October 24th 2010 has been used to explore the sensitivity of tau identification techniques in $t\bar{t}$ events. This analysis is elaborated in Chapter 4. The data collected between March and August 2011, with larger statistics than the 2010 data sample, are used to measure the top pair production cross section in the lepton plus tau channel. This measurement is detailed in Chapter

5.

The data taken with the ATLAS detector is separated into data periods in which trigger conditions and LHC operation mode remained stable. Each period consist of several segments of data, known as *luminosity blocks*. These blocks are included for the analysis if they were collected during periods in which the LHC was circulating stable colliding beams and all the ATLAS systems critical for $E_{\mathrm{T}}^{\mathrm{miss}}$ determination, muon, electron and jet identification were producing data with enough quality so as to be adequate for physics analysis [93]. Taking into account the luminosity block selection, the total integrated luminosity of the 2010 data sample is $35.3\mathrm{pb}^{-1}$, with an uncertainty of 3.2% [83]. For the 2011 sample the total integrated luminosity is calculated to be $2.05\ \mathrm{fb}^{-1}$. This luminosity estimate has an uncertainty of 3.7% [84, 85].

Monte Carlo simulation samples are used to optimize selection procedures, to calculate the signal acceptance and to evaluate the background contributions from single top quark, diboson (WW, WZ, ZZ) production, W+ jet events and $Z \to \tau^+\tau^-$ decays.

The simulation data is produced in three steps: generation of the event and immediate decays; simulation of the detector and physics interactions (using the GEANT4 [94] simulation of the ATLAS detector); and digitization, which is the process of converting GEANT4 simulated hits in active volumes of the detector to Raw Data Objects (RDOs), which act as input to the reconstruction package. The output of the simulation chain is presented in an identical format to the output of the ATLAS data acquisition system (DAQ). Thus, both the simulated and real data from the detector can then be run through the same ATLAS trigger and reconstruction packages [95].

For the $t\bar{t}$ signal and single top, the next-to-leading (NLO) generator MC@NLO [96], [97], [98] is used with a top quark mass of 172.5 GeV and with the NLO parton density function (PDF) set CTEQ66 [99]. The "diagram removal squeme" is used to remove single top and $t\bar{t}$ processes that produce identical final states at NLO [100, 101]. W + jets events with up to 5 partons and Z/γ^* + jets events with dilepton invariant mass $m_{\ell^+\ell^-} > 40$ GeV and up to 5 partons are generated by ALPGEN generator [102] using the MLM matching scheme [103] to remove overlaps between parton samples. W + jets, Z + jets and diboson background samples are generated using the LO PDF set CTEQ6L [99].

The cross-section of $t\bar{t}$ production is normalized to $164.57^{+11.45}_{-15.78}$ pb [27,28] [29] obtained from approximate next-to-next-to-leading-order (NNLO) perturbative Quamtum Chromodynamics (QCD) calculations. The cross-sections for W/Z + jets and diboson with jets have been rescaled by a factor of 1.2 to match NNLO calculations of their inclusive cross-sections.

The Alpgen W/Z + jets MC samples for 2011 analysis do not correctly reproduce the jet multiplicity in data. Scale factors (SF) derived by comparing event selections in data with the Alpgen MC have been applied. The values for the W + jets scale factors are listed in Table 3.1 and for $Z \to \mu\mu$ and $Z \to ee$ in Table 3.2.

In the case of application to $Z \to \tau\tau$, the scale factors for Z + jets are dependent on the p_{T} of the Z boson at generator level in order to ensure that the kinematic range of the $Z \to \mu\mu$ and $Z \to ee$ from which the scale factors are derived are similar to that of the $Z \to \tau\tau$ event that has a lower τ energy due to the neutrino decay. The derivation of the scale factors for Z + jets is documented in [104].

All samples that use HERWIG [105, 106] for parton shower evolution and hadronization rely on JIMMY [107] for the underlying event model. The τ lepton decays are handled by TAUOLA [108].

The effect of the multiple pp interactions per bunch crossing is taken into account in the MC

	$W \to \mu$ + jets		$W \to e$ + jets	
Jets	Pretag	Tagged	Pretag	Tagged
1	0.98 ± 0.11	0.95 ± 0.18	0.91 ± 0.11	0.85 ± 0.16
2	0.94 ± 0.10	0.91 ± 0.11	0.87 ± 0.09	0.81 ± 0.10
3	0.86 ± 0.10	0.83 ± 0.17	0.86 ± 0.15	0.80 ± 0.19
4	0.83 ± 0.10	0.80 ± 0.18	0.90 ± 0.12	0.83 ± 0.18
≥4	0.84 ± 0.12	0.82 ± 0.19	0.90 ± 0.13	0.83 ± 0.18
≥5	0.78 ± 0.11	0.75 ± 0.16	0.90 ± 0.21	0.84 ± 0.24

Table 3.1: *Scale factors applied to Alpgen W + jets MC for 2011 analysis.*

Jets	$Z(\mu\mu/ee)$
0	1.03 ± 0.002
1	0.96 ± 0.003
≥2	0.87 ± 0.005

Table 3.2: *Scale factors applied to Alpgen $Z \to \ell\ell$ MC for 2011 analysis.*

simulations in order to get a reasonable enough agreement with data to rely on the MC simulation for the signal acceptance calculation and part of the background estimations. As the event pileup increases, the number of detected particles will increase and it affects especially the missing transverse momentum and the jet multiplicity. The pileup effect is modelled by overlaying simulated minimum bias events over the original hard-scattering event [109]. MC events are then reweighted so that the distribution of interactions per crossing in the MC simulation matches that observed in data, which is $\langle\mu\rangle = 2$ for the 2010 dataset and $\langle\mu\rangle = 7$ for the 2011 data sample.

3.2 Physics Object selection criteria

The object selection for the analysis described below follows the official top working group recommendations for 2010 [110] and 2011 analysis [111] and is nearly the same object definition as in the $t\bar{t}$ cross section measurement in the dilepton channel [42], [43], with the exception of a τ candidate instead of a second electron or muon candidate.

3.2.1 Electrons

Electrons selected for the analysis are required, both in data and MC, to match an electron passing a single-electron trigger with a p_{T} threshold of 15 GeV for 2010 analysis or 20 GeV for 2011 analysis. The electrons that pass the trigger selection are then subjected to a set of quality requirements to ensure that candidates are consistent with the energy deposition (cluster) of an electron in the EM calorimeters and there is a well reconstructed track of charged particles in the inner detector associated with it and matching to the electromagnetic cluster. These electron candidates are required to pass stringent selection cuts on calorimeter, tracking and combined

variables that provide good separation between isolated electrons and jets. To further suppress the QCD multi-jet background, isolation requirements are applied.

The electron must have $p_{\mathrm{T}} > 20$ GeV (2010) or $p_{\mathrm{T}} > 25$ GeV (2011). Electrons also must satisfy $|\eta_{\mathrm{cluster}}| < 2.47$, excluding the barrel-endcap transition region ($1.37 < |\eta_{\mathrm{cluster}}| < 1.52$), where η_{cluster} is the pseudorapidity of the calorimeter energy cluster associated with the candidate. The ratio E/p_{T} between the electron cluster energy measured in the calorimeter and the momentum in the tracker must be consistent with that expected for an electron coming from a W boson.

Electrons from W boson decay are isolated from any jet activity. The three main sources of background for high momentum isolated electrons are hadrons faking an electron signature, electrons from heavy flavour decays and photon conversions. In order to suppress the background from these sources we require that there is little jet activity in the space surrounding the electron. The isolation criteria requires that the transverse energy deposited in the calorimeter in a cone of radius $\Delta R = 0.2$ around the electron candidate not associated with the electron must be less than 4 GeV (2010) or 3.5 GeV (2011).

The efficiency of the electron track and cluster reconstruction and the electron identification are measured within the Egamma Combined Performance Group using samples of $Z \to ee$ and $W \to e\nu$ in both data and MC for 2011 analysis [112]. Scale factors are derived to parametrise the different efficiency between data and MC.

3.2.2 Muons

The identification of muons in a candidate event is performed both at the trigger level and offline. Events with a possible muon candidate are selected at the trigger level requesting that a single-muon trigger with a p_{T} threshold of 18 GeV has been fired. The trigger request is applied to both collision data and MC simulated events.

Muon candidates are reconstructed offline by searching for track segments in different layers of the muon chambers. These segments are combined starting from the outermost layer, accounting for material effects, and matched with tracks found in the Inner Detector. The muon reconstruction combines an inner detector track with a muon spectrometer track using a global refit of the two tracks. The final candidates are refitted using the complete track information from both detector systems, and required to satisfy $p_{\mathrm{T}} > 4$ GeV and $|\eta| < 2.5$.

To suppress muons from heavy-flavour decays inside jets (such as b and c baryon and meson decays) and light flavour decays (π and K decays in flight) while favouring the $t \to W \to \mu\nu$ decay, muons must have a distance ΔR greater than 0.4 from any jet with $p_{\mathrm{T}} > 20$ GeV and low energy deposit in the calorimeter around the muon candidate. The sum of the p_{T} of the tracks around the muon candidate with $p_{\mathrm{T}} > 1$ GeV must also be low. The isolation requirements for 2010 analysis are:

- Calorimeter energy in a cone of $\Delta R = 0.2$ around the muon, $E_{\mathrm{T}}^{\mathrm{cone20}}$, less than 4 GeV
- The sum of track transverse momenta in a cone of $\Delta R = 0.2$ around the muon, $p_{\mathrm{T}}^{\mathrm{cone20}}$, less than 4 GeV

and for the 2011 analysis:

- Calorimeter energy in a cone of $\Delta R = 0.3$ around the muon, $E_{\mathrm{T}}^{\mathrm{cone30}}$, less than 4 GeV

- The sum of track transverse momenta in a cone of $\Delta R = 0.3$ around the muon, $p_{\mathrm{T}}^{\mathrm{cone30}}$, less than 4 GeV

The isolated muons are required to have $p_{\mathrm{T}} > 20$ GeV.

Muon candidates arising from cosmic rays are rejected by removing candidate pairs that are back-to-back in the $r - \phi$ plane and with transverse impact parameters relative to the beam axis $|d_0| > 0.5$ mm.

In 2011 MC samples, events where a muon candidate is found are applied a scale factor to account for the differences in the muon selection efficiency between data and MC. These scale factors are obtained by comparing the efficiency of the muon selection in data and MC $Z \to \mu\mu$ samples [112].

3.2.3 Jets

Jets are reconstructed with the anti-k_t [113, 114] algorithm (with distance parameter $R = 0.4$) starting from topological energy clusters made of adjacent cells in the EM calorimeters, reconstructed at the energy scale appropriate for electromagnetic objects (electrons and photons). The candidate jets are then calibrated to the hadronic energy scale, using a p_{T} and η dependent correction factor obtained from simulation [115].

Jet quality criteria are applied to identify the so-called bad jets as those not associated to in-time real energy deposits in the calorimeters caused by various sources ranging from hardware problems in the calorimeter, the LHC beam conditions, and the atmospheric muon-ray induced showers.

To avoid double counting, jets are removed if they overlap with an identified electron within $\Delta R < 0.2(0.4)$ in 2010 (2011) data. Only jet candidates with $p_{\mathrm{T}} > 20(25)$ GeV and $|\eta| < 2.5$ are used in 2010 (2011) analysis.

3.2.4 Missing Transverse Energy

The missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) is a measure of the energy imbalance in the transverse plane and is used as a signature to detect neutrinos.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ is constructed from the vector sum of all calorimeter cells with $|\eta| < 4.5$, calibrated at the electromagnetic scale (EM) and corrected according to the energy scale of the associated object. The results are then projected onto the transverse plane. Both high p_{T} and low p_{T} jets are considered. Energies of cells associated with jets are corrected using the jet energy scale. The τ jet candidates are treated as jets. Contributions from cells associated with electrons employ the electron energy calibration. Contributions from muons passing selection requirements are included. If the muon is not measured in the calorimeter, their momentum measured from the track inner detector and muon spectrometer system is used.

$$E_{\mathrm{x,y}}^{\mathrm{miss}} = E_{\mathrm{x,y}}^{\mathrm{Electrons}} + E_{\mathrm{x,y}}^{\mathrm{Jets}} + E_{\mathrm{x,y}}^{\mathrm{Muons}} + E_{\mathrm{x,y}}^{\mathrm{Other}}, \tag{3.1}$$

$$E_{\mathrm{T}}^{\mathrm{miss}} = \sqrt{\left(E_{\mathrm{x}}^{\mathrm{miss}}\right)^2 + \left(E_{\mathrm{y}}^{\mathrm{miss}}\right)^2}, \tag{3.2}$$

where $E_{\mathrm{x,y}}^{\mathrm{Other}}$ makes reference to the remaining clusters not associated with the high p_{T} objects (electrons, jets or muons) previously defined. More details can be found in references [116] and [117].

3.2.5 *b*-jets

One of the most important selection criteria for the analysis of events containing top quarks is the identification of jets containing *b*-quarks based on their specific properties: large *b*-hadron mass, long lifetime of *b*-hadrons and large branching ratio into leptons. The discrimination of *b*-jets from light quark jets is based in the relatively long lifetime of *b*-hadrons. This long lifetime results in a significant flight path length which leads to measurable secondary vertices, well separated from the primary interaction, and impact parameters of the decay products.

The key objects for *b*-jet reconstruction are tracks, jets and primary vertices. The accurate reconstruction of the charged tracks in the ATLAS Inner Detector is critical for a good *b*-tagging performance. The knowledge of the position of the primary interaction point, i.e., the primary vertex, of the proton-proton collision is important for *b*-tagging since it defines the reference point with respect to which impact parameters and vertex displacements are measured.

In the case of the 2010 data analysis, the SV0 tagger was used [118] [119]. SV0 is the simplest algorithm that exploits the secondary vertex information. It reconstructs secondary vertices from the tracks associated with a jet and returns the signed distance between the found secondary vertex and the primary vertex, divided by its error. A cut on the SV0 weight was set at 5.85 for a 50% signal efficiency.

In the case of the 2011 analysis, a secondary-vertex tagging algorithm, known as CombNN [120], is used. This algorithm constructs a likelihood ratio of *b*-jets and light-quark jets using the following discriminating variables: the signed impact parameter significance of well measured tracks associated with a given jet, the decay length significance associated with a reconstructed secondary vertex, the invariant mass of all tracks associated with the secondary vertex, the ratio of the sum of the energies of the tracks associated with the secondary vertex to the sum of the energies of all tracks in the jet assuming a pion hypothesis, and the number of two-track vertices that can be formed at the secondary vertex.

The cut on the combined likelihood ratio has been set at 0.35 to give an average efficiency of 70% for identifying a *b*-jet and a 1% efficiency for light-quark and gluon jets[1]. The rejection factor is estimated to be 5 for charm jets and 99 for light flavor jets.

The performance of the CombNN *b*-tagging algorithm in data and MC shows some discrepancies. Tagging and mis-tagging efficiency scale factors correct these discrepancies and are estimated based on *b*-tagging calibration studies. Scale factors depend on the p_{T} and flavour (*b*, *c*, or light flavour) of the jets.

3.2.6 Reconstructed tau leptons

Tau leptons produced in pp collisions decay before reaching the detector. Therefore, tau leptons are identified via their visible decay products, which are summarized in Figure 3.1.

About 35% of the time, τ leptons decay leptonically:

$$\tau \to \nu_\tau + \nu_e + e \qquad (17.4\%)$$

$$\tau \to \nu_\tau + \nu_\mu + \mu \qquad (17.8\%)$$

[1] Jets originating from the hadronization of light-quarks and gluons are referred to in the text as light-quark jets and gluon jets, respectively.

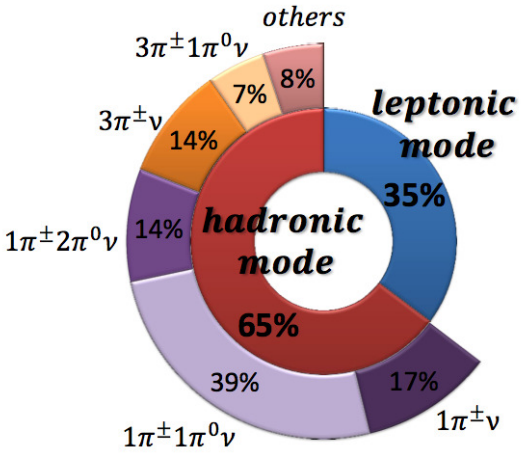


Figure 3.1: *Tau lepton decaying modes.*

and the remaining 65% of the time they decay hadronically, producing very collimated jets of charged and neutral pions, called τ jets. Hadronic decays to only one charged pion are called "*single prong*" (1p) decays:

$$\tau \rightarrow \pi^{\pm}\nu \qquad (11.0\%)$$

$$\tau \rightarrow n\pi^{0}\pi^{\pm}\nu \qquad (37.6\%)$$

and hadronic decays to three charged pions are called "*multi prong*" (mp) decays:

$$\tau \rightarrow 3\pi^{\pm}\nu \qquad (9.82\%)$$

$$\tau \rightarrow n\pi^{0}3\pi^{\pm}\nu \qquad (3.89\%)$$

There is a small "*five prong*" contribution which is really hard to detect in a jet environment and it is not considered in this analysis. Single-prong and multi-prong tau decays will be referred as τ_1 and τ_3, respectively.

To distinguish at detector level between leptons coming directly from W bosons or those from the tau decay is extremely complicated, since their signal is identical. Thus, we focus in the reconstruction and identification of hadronically decaying tau leptons.

The reconstruction and identification of hadronically decaying tau leptons proceeds as follows:

1. the tau reconstruction starts by considering each reconstructed jet as a τ candidate. Jets are reconstructed using the anti-k_t algorithm, with distance parameter R = 0.4, starting from topological energy clusters[2] in the calorimeter.

[2]Collection of neighbouring cells for which the signal is significantly above the noise threshold.

2. Information from the calorimeter and the Inner Detector is used to search for key aspects of tau jets. The Inner Detector provides information on the charged hadronic 1 track or 3 track system, whereas the calorimeter provides information on the energy deposition from the visible decay products:

   - Reconstruction and identification of "good quality" tracks: $p_{\mathrm{T}} > 1$ GeV, minimum number of hits in all Inner Detector components (Pixel, SCT and TRT), transverse impact parameter $|d_0| < 1$ mm and good χ^2 on the fit for the trajectory reconstruction. All good tracks with $p_{\mathrm{T}} > 1$ GeV inside a cone of $\Delta R < 0.4$ around the jet axis are associated to the τ candidate.
   - Apply a veto against muon tracks: tracks should not be matched to track segments in the muon spectrometer.
   - Apply a veto against electron tracks, which show high threshold hits in the TRT.
   - Identification of at least a high p_{T} leading hadronic track ($p_{\mathrm{T}} > 4$ GeV) that passes the quality requirements. In the 1p case we require no nearby tracks. For the 3p one or two more tracks are required around the leading track. Cases with only two good quality tracks identified are considered as 3p, when the third track is present in the event, but not passing the quality criteria.
   - Since tau jets are highly collimated, all tracks are searched for in a cone of $\Delta R = 0.2$ around the leading track. An outside region with $0.2 < \Delta R < 0.4$, called the isolation cone, is used to define isolation criteria. An schematic view of the tau jet tracks with the signal and isolation cones is shown in Figure 3.2.
   - The reconstructed τ charge is measured from the charged pion tracks. Only reconstructed tau leptons with charge $\neq 0$ are considered in the analysis.
   - Collected calorimeter energy deposition in a fixed cone around the leading track is used to calculate kinematic quantities (such as E_{T}) of the τ candidate.

3. Since hadronic τ decays consist of a mix of charged and neutral pions, they produce a deposition of energy in the hadronic calorimeter from the charged pions and in the electromagnetic calorimeter from the photons originated in the π^0 decays. Therefore, the energy of the tau jets has to be calibrated both at EM and hadronic level. The final tau energy scale is determined from MC studies, comparing the true visible energy of the tau to the simulated detector response. The details of the tau energy scale determination can be found in [121].

Details of the tau lepton reconstruction are given in [122] for 2010 analysis and in references [123] and [124] for 2011 analysis, respectively.

The reconstructed τ candidates used for analysis must have $p_{\mathrm{T}}^{\tau} > 15(20)$ GeV in 2010(2011) data, $|\eta_{\tau}| < 2.3$ and one, two or three associated tracks ($0< N_{\mathrm{trk}} <4$). The lead τ track must have $p_{\mathrm{T}} > 4$ GeV. The τ charge is given by the sum of the charges of the associated tracks, and is required to be non-zero. The probability of misidentifying the τ lepton charge sign is about 1%. The charge misidentification rate for muons and electrons is negligible. Additional requirements are made to ensure the τ candidates must have been found by the calorimeter-based algorithm.

Tau candidates overlapping with non isolated muons with $p_{\mathrm{T}}^{\mu} > 4$ GeV and electron candidates with $p_{\mathrm{T}} > 15$ GeV within a cone of radius $\Delta R(\ell, \tau) < 0.4$, are removed. τ candidates overlapping

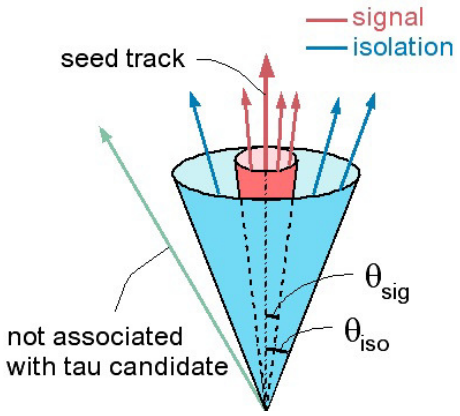


Figure 3.2: *Tau isolation cone.*

with b-tagged jets within a cone of radius $\Delta R(b\text{-jet}, \tau) < 0.4$ are also removed. To remove electrons identified as τ leptons that have not have been identified as a tight electron candidate, we apply a medium Boosted Decision Tree (BDT_e) based electron veto. The Boosted Decision Tree technique is described in Section 3.3.3.

The loose τ candidates used in this analysis are required to have $BDT_e > 0.5\,(0.51)$ in 2010 (2011) data. This cut has an efficiency of 85% for hadronically decaying taus in $Z \to \tau^+\tau^-$ events. The additional rejection for electrons, after removing overlap of reconstructed electrons with τ candidates, is a factor of 60.

The majority of objects reconstructed as τ candidates in a multi-jet environment are jets misidentified as τ leptons. A jet or an electron misidentified as a τ lepton will be referred to as a fake τ. In Section 3.3, τ identification techniques to differentiate taus from electrons and other jet types are detailed.

The fake τ background in the τ_3 sample is significantly higher than in the τ_1 sample, as we will see in Sections 4.1 and 5.2. Furthermore, the variables used for τ identification exhibit different behaviour for single-prong and multi-prong τ candidates. Hence, independent measurements are done for τ_1 and τ_3 candidates.

3.3 Tau Identification Techniques

The main source of misidentified τ leptons is the QCD multi-jet background, which has a production cross section many orders of magnitude above the cross sections for weak interaction processes involving tau leptons.

Jets are showers of particles produced through a process called hadronization. After high-energy collisions in a particle collider, in which free quarks or gluons are created, gluons spontaneously combine with quarks and antiquarks producing stable hadrons, mainly through qq, qg, and gg interactions. The result is a stream of collinear hadrons with total momentum approximately the same as the original outgoing quark. A jet can contain any hadron kinematically allowed, but it is dominated by charged $\pi^{\pm}$ and neutral pions π^0.

The identification of hadronically decaying τ leptons is based on a series of identification variables that exploit the differences between τ and QCD jets topology. The tau jets are characterized by:

- A collimated jet, with 90% of energy in a cone of $\Delta R = 0.2$ around leading track. Therefore, they present a narrow shower in electromagnetic calorimeter.
- Low track multiplicity (1-3 charged tracks associated with charged pions).
- The leading track carries most of the jet energy.
- A lower average π^0 multiplicity than other jet types, thus having a higher deposition in the electromagnetic calorimeter.
- A low invariant mass of tracks and clusters.
- They have no gluon radiation, therefore isolation from the rest of the event is required both in the Inner Detector and the Calorimeter.

Electrons can also be misidentified as 1-prong τ leptons. Separate procedures must be developed for rejecting electrons as their signature is different from QCD jets. The following differencies between τ and electron signatures in the detector are used to build a multivariate discriminant, BDT_e, to remove electrons reconstructed as taus:

- The shower produced by a tau lepton in the calorimeter tends to be longer and wider than an electron-induced shower.
- Electrons produce a larger signal in the TRT due to their lower mass and higher gamma[3].

Tau identification takes advantage of these distinguishing characteristics to built a set of discriminant variables (defined in Section 3.3.1). Two different approaches are then used for τ identification. The first one applies a set of optimized cuts on a reduced set of variables. It is a more conservative method, known as the Safe Cuts (SC) method and it is described in section 3.3.2. On the second approach, the identification variables are combined into multivariate discriminants using a Boosted Decision Tree technique and the outputs of the discriminants are used to separate jets and electrons misidentified as τ leptons decaying hadronically from real τ leptons. This method is explained in more detail in section 3.3.3.

Details on the variables definitions and both the safe cuts and BDT identification methods are given in reference [122] for the 2010 analysis and in references [123] and [124] for the BDT optimization to be used in the 2011 data.

3.3.1 Tau Identification variables

To separate real τ jets from background several discriminant variables are used, based on track and calorimeter information. These variables, defined below, exploit the critical aspects that differentiate tau jets from electrons and other jet types:

[3]The probability of emitting transition radiation depends on a particle's relativistic gamma factor (E/m).

1. **Transverse momentum fraction of the leading track**:

$$f_{trk,1} = \frac{p_{\mathrm{T}}(lead\ track)}{p_{\mathrm{T}}(\tau_{had})}. \tag{3.3}$$

The leading track in a τ jet carries a high fraction of its energy, while QCD jets are expected to have a more uniform distribution of p_{T} among their tracks. Therefore, $f_{trk,1}$ will have higher values for τ jets.

2. **Electromagnetic radius**. The energy weighted shower width in the EM calorimeter:

$$R_{EM} = \frac{\sum_{i=1}^{n} \Delta R_i \cdot E_{\mathrm{T}i}}{\sum_{i=1}^{n} E_{\mathrm{T}i}}, \tag{3.4}$$

where i runs over all electromagnetic calorimeter cells in the cluster with $\Delta R < 0.4$ around the jet axis, n is the number of cells and $E_{\mathrm{T}i}$ is the transverse energy in cell i. The electromagnetic radius allows one to exploit the small transversal shower profile of τ lepton decays in the electromagnetic calorimeter.

3. **Track radius**. Defined in the same way as R_{EM}, except that track p_{T} is used instead of calorimeter cell E_{T} in the weighting:

$$R_{trk} = \frac{\sum_{i=1}^{n} \Delta R_i \cdot P_{Ti}}{\sum_{i=1}^{n} P_{Ti}}. \tag{3.5}$$

All tracks associated to the τ candidate within $\Delta R < 0.2$ are considered.

4. **Electromagnetic Fraction** is the ratio of the calibrated E_{T} in the electromagnetic calorimeter to the total calibrated electromagnetic and hadronic E_{T}:

$$F_{EM} = \frac{E_{T_{EM}}^{calib}}{E_{T_{EM}}^{calib} + E_{T_{Had}}^{calib}}. \tag{3.6}$$

For τ jets this variable depends highly on the number of neutral pions produced in the decay. On average, τ leptons deposit a larger fraction of their energy in the EM calorimeter than other jet types.

5. **Centrality Fraction** is a measure of how concentrated calorimeter energy deposits are with respect to the jet axis. Energy deposits from τ jets tend to be more concentrated, with higher centrality fraction values, than QCD jets, but less concentrated than electrons.

$$F_{core} = \frac{\sum_i E_{T_i}}{\sum_j E_{T_j}}, \tag{3.7}$$

where i runs over all electromagnetic calorimeter cells in the cluster associated with the τ candidate with $\Delta R < 0.1$ and j runs over cells within $\Delta R < 0.4$.

6. **Invariant mass of the topological clusters of calorimeter cells**:

$$M_{topo} = \sqrt{\left(\sum_{\text{clusters}} E\right)^2 - \left(\sum_{\text{clusters}} \mathbf{p}\right)^2}. \tag{3.8}$$

This variable tends to peak near the τ invariant mass for real τ leptons and it is relatively unconstrained in QCD jets.

Two more variables are used to identify τ_3 candidates:

7. **Invariant mass of the tracks**: constrained by the actual τ mass for τ jets but scaled with the energy of a QCD jet. This variable definition is valid for τ candidates with two or more tracks.

$$M_{trk} = \sqrt{\left(\sum_{\text{tracks}} E\right)^2 - \left(\sum_{\text{tracks}} \mathbf{p}\right)^2} \tag{3.9}$$

8. **Transverse Flight Path significance**: the transverse displacement of the secondary vertex from the primary one divided by its uncertainty. The secondary or τ decay vertex is defined (for τ candidates with multiple reconstructed tracks) as the common origin of all charged tracks. Tau leptons have a significant lifetime, therefore, they are expected to have a longer flight length than typical QCD jets. The tau-decay vertex is separated from its production point, whereas secondary vertices from QCD jets have a larger overlap with the primary vertex.

One last variable, not used for jet rejection:

- **Number of high threshold hits over the number of low threshold hits in the TRT**: since electrons produce a larger signal in the TRT due to their lower mass and higher gamma, this variable is the most powerful to discriminate τ candidates from electrons.

3.3.2 Cut-Based Tau Identification

In early 2010 data, the τ identification was based on a simple cut-based approach using three of the previously defined variables: the Electromagnetic radius (R_{EM}), track radius (R_{trk}) and transverse momentum fraction of the leading track ($f_{trk,1}$). R_{EM} and R_{trk} quantify the width of the hadronic shower, which tends to be larger for QCD jets than for τ leptons of the same energy. These variables were selected since they were robust and well modelled in early ATLAS data.

The R_{EM} and R_{trk} cuts were parametrized as a function of p_{T} of the τ candidate to remove the p_{T} dependence from the identification efficiency. The optimization procedure for choosing the cut values is documented in [125]. Cut combinations were selected for 30% (loose), 50% (medium) and 60% (tight) signal efficiency in the selection of $Z \to \tau\tau$ events. These cuts are shown in Tables 3.3 and 3.4 for τ_1 candidates and Tables 3.5 and 3.6 for τ_3 candidates.

3.3.3 Tau Identification using Boosted Decision Trees

The Boosted Decision Trees (BDT) Tau ID is described in detail in [126]. This method makes use of all the variables defined in 3.3.1 as input to a multi-variable technique which produces a

	R_{EM}	R_{trk}	$1/f_{trk,1}$
Loose Cut	$0.063 - 2.04e^{-4} \cdot p_{\mathrm{T}} + 0.663/p_{\mathrm{T}}$	$0.134 - 2.72e^{-4} \cdot p_{\mathrm{T}} + 0.0328/p_{\mathrm{T}}$	-
Medium Cut	$0.0522 - 1.71e^{-4} \cdot p_{\mathrm{T}} + 0.715/p_{\mathrm{T}}$	$0.134 - 2.72e^{-4} \cdot p_{\mathrm{T}} + 0.0328/p_{\mathrm{T}}$	8.33
Tight Cut	$0.0306 - 1.03e^{-4} \cdot p_{\mathrm{T}} + 0.819/p_{\mathrm{T}}$	$0.134 - 2.72e^{-4} \cdot p_{\mathrm{T}} + 0.0328/p_{\mathrm{T}}$	7.14

Table 3.3: *Criteria for Safe Cuts Tau ID for τ_1 candidates with $p_{\mathrm{T}} < 80$ GeV.*

	R_{EM}	R_{trk}	$1/f_{trk,1}$
Loose Cut	0.055	0.113	-
Medium Cut	0.0475	0.113	8.33
Tight Cut	0.0325	0.113	7.14

Table 3.4: *Criteria for Safe Cuts Tau ID for τ_1 candidates with $p_{\mathrm{T}} \geq 80$ GeV.*

	R_{EM}	R_{trk}	$1/f_{trk,1}$
Loose Cut	$0.179 - 5.17e^{-4} \cdot p_{\mathrm{T}} + 0.339/p_{\mathrm{T}}$	$0.0565 - 1.44e^{-4} \cdot p_{\mathrm{T}} + 0.695/p_{\mathrm{T}}$	4.55
Medium Cut	$0.162 - 4.62e^{-4} \cdot p_{\mathrm{T}} + 0.447/p_{\mathrm{T}}$	$0.0303 - 9.9e^{-5} \cdot p_{\mathrm{T}} + 0.810/p_{\mathrm{T}}$	3.33
Tight Cut	$0.0833 - 2.15e^{-4} \cdot p_{\mathrm{T}} + 0.930/p_{\mathrm{T}}$	$0.0146 - 7.21e^{-5} \cdot p_{\mathrm{T}} + 0.879/p_{\mathrm{T}}$	2.5

Table 3.5: *Criteria for Safe Cuts Tau ID for τ_3 candidates with $p_{\mathrm{T}} < 80$ GeV.*

	R_{EM}	R_{trk}	$1/f_{trk,1}$
Loose Cut	0.142	0.0536	4.55
Medium Cut	0.13	0.0325	3.33
Tight Cut	0.0777	0.0198	2.5

Table 3.6: *Criteria for Safe Cuts Tau ID for τ_3 candidates with $p_{\mathrm{T}} \geq 80$ GeV.*

continuous discriminant output that later on is used to measure to what degree the τ candidate resembles a real tau jet.

The power of the BDT method relies on its higher discriminant power when compared to the individual variables used as input. In simple cut-based techniques, any candidate which fails the first cut is immediately discarded from the analysis sample. When using Decision Trees, candidates which fail a single cut are not immediately discarded from the sample, but are re-examined by the algorithm using all remaining variables. It is a multiple cut technique that allows to recover the signal candidates which would otherwise be lost in a simple cut method and remove background that would normally pass.

Decision trees are first constructed or "trained" using a sample of known signal and background composition called the training sample:

- The first step selects the variable and optimal cut value which provides the best signal to background separation and divide the sample in two subsets: those events that pass the cut and those that do not pass it.

- Then this process is repeated iteratively with the remaining variables until all have been used, or all subsets are pure signal or pure background or the number of events is too few to continue.
- Each decision point in the tree in which a variable and cut value are provided and the candidate is determined to either pass or fail it is called a node.
- At the end, if a subset has purity greater than 1/2 (or any defined cut value), then it is labelled as signal and if the purity is less than 1/2, it is a labelled as background. Events are classified signal if they land on a signal set and background if they land on a background set. The resulting tree is a decision tree and the ending data set are their leaves.
- For each event, the decision tree result (or classifier value) D_i is equal to the purity of the leaf on which the testing terminates. The more signal-like an event is, the closer its classifier value is to 1.

The key element of the BDT training process is therefore the criterion used to determine the optimal cut for each variable at each node. BDT training for τ identification uses the Gini index, defined as

$$i_{GINI} = p(1 - p), \tag{3.10}$$

where p is the signal purity at each node and (1-p) is the background purity. The purity of a sample at a given node is defined as

$$p = \frac{\sum_s w_s}{\sum_s w_s + \sum_b w_b}, \tag{3.11}$$

where $\sum_s w_s$ and $\sum_b w_b$ are the weighted sums of the signal and background events contained within the node. Before training begins, events are weighted by the cross section for the sample divided by the number of events in the sample. Thus, the Gini index can be written as

$$i_{GINI} = p(1 - p) = \frac{\sum_s w_s \ \sum_b w_b}{(\sum_s w_s + \sum_b w_b)^2}. \tag{3.12}$$

The Gini index has a minimum of zero that occurs if $p = 1$ (all signal) or $p = 0$ (all background). The optimal cut minimizes the i_{GINI} index at each node, therefore maximizing the background purity, 1-p, on one side of the cut and maximizing the signal purity, p, on the other side of the cut. The optimization procedure to find the appropriate variable and cut value is well documented in [127].

Decision trees are a powerful but unstable tool to separate signal from background events. Small changes in the training sample can give a large change in the tree and the results. Also, as the number of nodes increases, the number of events available in each node decreases and statistical fluctuations can affect the decision procedure and produce the tree to branch in a way that does not truly correspond to the object under study.

A common way to address this weakness is through averaging several decision trees in order to determine the final probability that a candidate is signal. In addition to improving stability, the process generally improves performance, providing a better signal to background ratio of the sample. The averaging technique used for τ identification, known as "boosting", uses the adaptive boosting algorithm AdaBoost [128], and proceed as follows:

- After passing a decision tree, if a training event with weight w_i is misclassified, i.e, a signal event lands on a background leaf or a background event lands on a signal leaf, then the weight of that event is increased (boosted), in order to increase their importance during the next training. The weights of correctly classified candidates are not changed.

- The boosting parameter used to reweight the misclassified objects for the ith decision tree is defined by

$$\alpha_i = \beta \ln\left(\frac{1-\epsilon_i}{\epsilon_i}\right), \tag{3.13}$$

  where ϵ_i is the error fraction of the ith decision tree which is the sum of the weights of the misclassified event over the sum of the weights of all objects:

$$\epsilon_i = \frac{\sum w_{i,\text{misclasified}}}{\sum w_i}, \tag{3.14}$$

  and β is a parameter set by the user which scales the amount by which the weights are boosted/suppressed in the next tree. For τ identification it is set at $\beta = 0.2$.

- The new training sample with the new weights is used to build a new decision tree, in which the order of the variables and the optimized cut values may differ from the previous one. The new tree focus more on difficult cases and less on easy cases.

- The process is reiterated until no events in the training sample are misclassified.

The final classifier value, or BDT score, of an event after passing through the Boosted Decision Tree is the weighted sum of the individual tree results D_i:

$$BDT_{\text{score}} = \sum_{i}^{N_{trees}} \alpha_i \cdot D_i. \tag{3.15}$$

High scores mean the object is most likely signal and low scores that it is most likely background. By choosing a particular value of the score on which to cut, one can select a desired fraction of the signal or a desired signal to background ratio.

Boosted Decision Trees are widely used in high energy physics, and have already shown their effectiveness, such as in the first evidence for single top quark production analysis by the DØ Collaboration [129].

For τ lepton identification, the BDT discriminant was trained using jets from data as background and τ leptons from $Z \to \tau\tau$ MC as signal. For each τ candidate a BDT score (BDT_j) is calculated that has values that are more closer to 1.0 the more the candidate is like a real τ, and closer to 0.0 the less the candidate resembles a τ.

The cut in the BDT_j has to be such that it maximizes the acceptance on real τ leptons while minimizing that of false τ jets. The optimized cut value for 2010 analysis was found to be $BDT_j > 0.6$ and it was increased to $BDT_j > 0.7$ for 2011 analysis.

To remove electrons identified as τ leptons that may not have been identified as an electron we use an additional criteria, relying on a Boosted Decision Tree output (BDT_e) trained to separate τ leptons and electrons, using simulated $Z \to \tau^+\tau^-$ samples as signal and $Z \to e^+e^-$ MC as

background. The most effective variables to construct the BDT_e are E/p, the EM fraction and the cluster-based shower width. As with any other BDT output, BDT_e tends to be near 1 (0) if the τ candidate is a τ lepton (electron). The medium working point corresponds to 85% efficiency for $Z \to \tau\tau$. This is an official veto supported by the ATLAS Tau Working Group which relies on a BDT trained to separate taus and electrons and has been documented in [123].

4

Tau identification studies with early ATLAS data.

In the Standard Model the top quark decays almost 100% of the times to a W boson and a b quark. Signal events produced from top quark pair production in the lepton plus tau channel:

$$t\bar{t} \rightarrow W(\rightarrow e/\mu + \nu_{e/\mu})bW(\rightarrow \tau + \nu_{\tau})b \tag{4.1}$$

will exhibit the following experimental final state topology in the ATLAS detector (see Figure 4.1): two jets with high transverse energy E_{T}, one isolated lepton (electron or muon) with high transverse momentum (p_{T}) from one W boson decay, one hadronically decaying τ from the other W boson decay and large missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) from the undetected neutrinos.

The backgrounds affecting the $t\bar{t}(\ell, \tau)$ signal can be classified in three types:

1. $t\bar{t}$ dileptonic and semileptonic decays where a jet or a lepton has been falsely identified as a tau jet
   - $t\bar{t} \rightarrow W(\rightarrow e/\mu + \nu_{e/\mu})bW(\rightarrow qq^{'})b$
   - $t\bar{t} \rightarrow W(\rightarrow e/\mu + \nu_{e/\mu})bW(\rightarrow e/\mu + \nu_{e/\mu})b$
2. Non $t\bar{t}$ events with real tau leptons
   - $Z \rightarrow \tau^{+}\tau^{-}$ events.
   - Diboson production WW, WZ, ZZ, where one boson decays into a lepton and the other into a tau.
   - Single top, mainly from the $Wt(\rightarrow Wb)$ production channel
3. Non $t\bar{t}$ events with a jet misidentified as a tau candidate

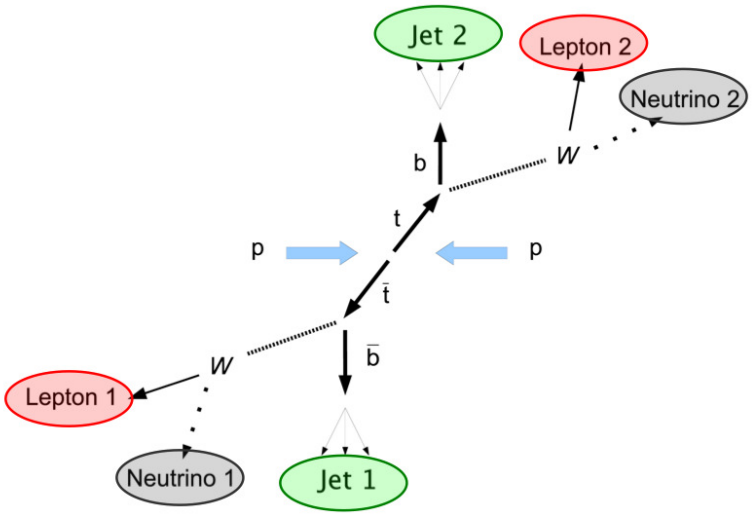


Figure 4.1: *Scheme of a $t\bar{t}$ interaction decaying into two leptons. The final state in the detector is characterized by two b-jets with high transverse energy E_{T}, two isolated leptons (one of which will be an hadronically decaying τ in the $t\bar{t} \to \ell + \tau$ channel) and large missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) from the undetected neutrinos.*

- W+ jet events
- Multi-jet events where one jet has been identified as a lepton and the other one as a tau jet.

The selection of $t\bar{t} \to \ell + \tau$ events, described in Section 4.1, is based on kinematical requirements optimized for the 2010 data samples. After applying the event selection, the dominant backgrounds are W+ jet and top pair production in the lepton plus jets channel. After applying b-tagging $t\bar{t}$ lepton plus jets becomes the dominant background. Therefore, τ lepton identification is the critical discriminator between signal and background. The two different identification criteria described in Section 3.3, safe cuts and Boosted Decision Trees, are studied using 35 pb^{-1} of integrated luminosity data collected in 2010 to explore the feasibility of $t\bar{t} \to \ell + \tau$ measurements. The $t\bar{t} \to \ell + \tau$ background from misidentified jets is estimated from data using a simple matrix method technique described in Section 4.3.

4.1 $t\bar{t} \to \ell + \tau$ event selection

The following event selection has been optimized to maximize the signal significance using Monte Carlo samples:

- Data event cleaning is needed in order to ensure the quality of object reconstruction. The Data Quality (DQ) selection within ATLAS is based upon inspection of a standard set of distributions that leads to a Data Quality assessment which is encoded in the so-called

DQ flags. DQ flags are issued for each detector, trigger, and for each physics object (jets, electrons, muons, etc.) to be reconstructed. Hence the state of the ATLAS detector, from hardware to physics object reconstruction, is expressed through the DQ flags, which are saved per luminosity block [130]. Luminosity blocks that pass the DQ selections are included in a Good Run List (GRL). Events not included in the GRL are discarded from the analysis.

- In order to ensure the event originated from the collision and to reject the non-collision background events, an event is required to have a primary vertex with at least five tracks and is discarded if any jet with $p_{\mathrm{T}} > 20$ GeV fails the jet quality selections designed to reject jets arising from out-of-time activity or calorimeter noise [130].

- There must be one, and only one, isolated electron (muon) and no identified muons (electrons) for the $e(\mu) + \tau$ channel. This requirement vastly reduces the multi-jet background.

- The selected isolated lepton must be matched to a single lepton trigger: a single-electron trigger with a p_{T} threshold of 15 GeV or a single-muon trigger with a p_{T} threshold between 10 and 13 GeV .

- There must be at least one loose τ candidate, as defined in Section 3.2.

- There must be at least two jets not overlapping with a τ candidate. In the case of two τ candidates, we chose τ_1 over τ_3 as a τ candidate. If both are either τ_1 or τ_3, at least one jet is required to be not overlapping with either τ candidate, and both candidates are kept.

- A cut on the transverse missing energy, $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV, is applied to reduce the QCD multi-jet background. Multi-jet events have low $E_{\mathrm{T}}^{\mathrm{miss}}$ since the only neutrinos produced come from secondary decays inside the jets.

- A cut on the scalar sum of the lepton p_{T}, the p_{T} of the jets, the p_{T} of the τ candidate, and $E_{\mathrm{T}}^{\mathrm{miss}}$, $H_{\mathrm{T}} > 200$ GeV, is applied to reduce the W + jets background.

- There must be at least one jet identified as b-jet (≥ 1 b-tag) by applying the b-tagging requirement SV0 weight ≥ 5.85. The efficiency of this cut is 50% for b-jets from $t\bar{t}$ decays.

cut	$t\bar{t}(\ell,\tau)$	$t\bar{t}(\ell$+jets)	$t\bar{t}(\ell\ell')$	Z+jets	W+jets	$t(\to Wb)q$	Total	Data
Isolated e	49 ± 1	442 ± 2	83 ± 1	15627 ± 39	133705 ± 104	16 ± 0	149922 ± 111	320346
Trigger	48 ± 1	441 ± 2	82 ± 1	15514 ± 39	132806 ± 104	16 ± 0	148907 ± 111	315910
≥ 1 τ candidate	23 ± 0	152 ± 1	17 ± 0	1333 ± 11	6378 ± 20	2 ± 0	7906 ± 23	27874
$N_{\mathrm{jet}} \geq 2$	20 ± 0	148 ± 1	14 ± 0	281 ± 5	841 ± 5	1 ± 0	1305 ± 7	3360
$E_T^{\mathrm{miss}} > 30$ GeV	17 ± 0	118 ± 1	12 ± 0	75 ± 3	567 ± 4	1 ± 0	791 ± 5	1069
$\sum E_T > 200$ GeV	17 ± 0	117 ± 1	12 ± 0	60 ± 2	391 ± 3	1 ± 0	598 ± 4	708
≥ 1 bjet	12 ± 0	79 ± 1	8 ± 0	2 ± 0	18 ± 1	0 ± 0	120 ± 1	166

Table 4.1: *$e + \tau_1$ Cut Flow. $t\bar{t}(\ell,\tau)$ are the expected signal events. $t\bar{t}(\ell\ell')$ are $t\bar{t}$ dilepton events with only one lepton reconstructed as a lepton and a lepton reconstructed as a τ. $t\bar{t}(\ell$+jets) are semileptonic events with one lepton and a jet reconstructed as a τ. After the b-jet requirement $t\bar{t} \to \ell$ + jet is the dominant background.*

The number of MC and data events passing each step of the event selection are shown in Table 4.1 for τ_1 candidates and Table 4.2 for τ_3 candidates for the $e + \tau$ channel. The excess of data over prediction is attributed to the QCD multi-jet background, which is in not included in the MC simulation since QCD MC suffers from low statistics (especially in the $e + \tau$ channel) and the detector simulation is insufficiently precise to model the rate of jets falsifying isolated leptons (electrons or muons) and tau candidates.

cut	$t\bar{t}(\ell,\tau)$	$t\bar{t}(\ell$+jets)	$t\bar{t}(\ell\ell')$	Z+jets	W+jets	$t(\rightarrow Wb)q$	Total	Data
Isolated e	49 ± 1	442 ± 2	83 ± 1	15627 ± 39	133705 ± 104	16 ± 0	149922 ± 111	320346
Trigger	48 ± 1	441 ± 2	82 ± 1	15514 ± 39	132806 ± 104	16 ± 0	148907 ± 111	315910
≥ 1 τ candidate	14 ± 0	189 ± 1	29 ± 0	1189 ± 11	8930 ± 24	5 ± 0	10356 ± 26	46406
$N_{jet} \geq 2$	11 ± 0	179 ± 1	21 ± 0	289 ± 5	903 ± 5	1 ± 0	1404 ± 7	3817
$E_T^{miss} > 30$ GeV	9 ± 0	142 ± 1	19 ± 0	67 ± 2	612 ± 4	1 ± 0	850 ± 5	1273
$\sum E_T > 200$ GeV	9 ± 0	141 ± 1	18 ± 0	55 ± 2	456 ± 3	1 ± 0	680 ± 4	917
≥ 1 bjet	7 ± 0	100 ± 1	13 ± 0	2 ± 0	26 ± 1	1 ± 0	148 ± 1	237

Table 4.2: *$e + \tau_3$ Cut Flow. $t\bar{t}(\ell,\tau)$ are the expected signal events. $t\bar{t}(\ell\ell')$ are $t\bar{t}$ dilepton events with only one lepton reconstructed as a lepton and a lepton reconstructed as a τ. $t\bar{t}(\ell$+jets) are semileptonic events with one lepton and a jet reconstructed as a τ. After the b-jet requirement $t\bar{t} \rightarrow \ell$ + jet is the dominant background.*

Figure 4.2 shows the lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ transverse mass distribution (M_t) before applying the b-tagging requirement. The agreement between data and MC is quite good at high M_t, while at low M_t there is a significant excess in data. Since the QCD multi-jet background is expected to have lower M_t values, the disagreement between data and MC, which is is larger at the first steps of the event selection and gets smaller from the selection of at least one tau candidate onward, can be safely attributed to the multi-jet background.

After applying the event selections, the dominant background is W+ jets, followed by $t\bar{t} \rightarrow \ell$ + jet. The b-jet identification highly suppresses the W+ jets background. The dominant background after all event selections and b-tagging are applied is $t\bar{t} \rightarrow \ell$ + jet, where a jet has been misidentified as a τ candidate. The $t\bar{t}$ semileptonic background constitutes 66% of the expected background for τ_1 and 68% for τ_3.

At this stage, the only discerning tool between signal and background is τ identification.

4.2 Tau identification methods

Tau identification is based on a short set of variables that were robust and well modelled with early ATLAS data. These variables were described in Section 3.3.1.

The safe cuts method applies energy dependent cuts on three variables: electromagnetic radius (R_{EM}), track radius (R_{trk}) and the transverse momentum fraction of the leading track (f_{trk}). The optimized cut values were detailed in Section 3.3.2. In this analysis we are using medium cuts for τ_1 and tight cuts for τ_3, which are summarized in Table 4.3.

Figure 4.3 shows the three variables used for the cut-based tau identification after applying preselection and before b-tagging for τ_1 and τ_3 candidates, respectively.

A Boosted Decision Tree based τ identification, described in Section 3.3.3, is also applied to the b-tagged data and MC samples. The cut in the BDT_j variable has to be such that it maximizes the acceptance on real τ leptons while minimizing that of false τ jets. The BDT_j distributions for

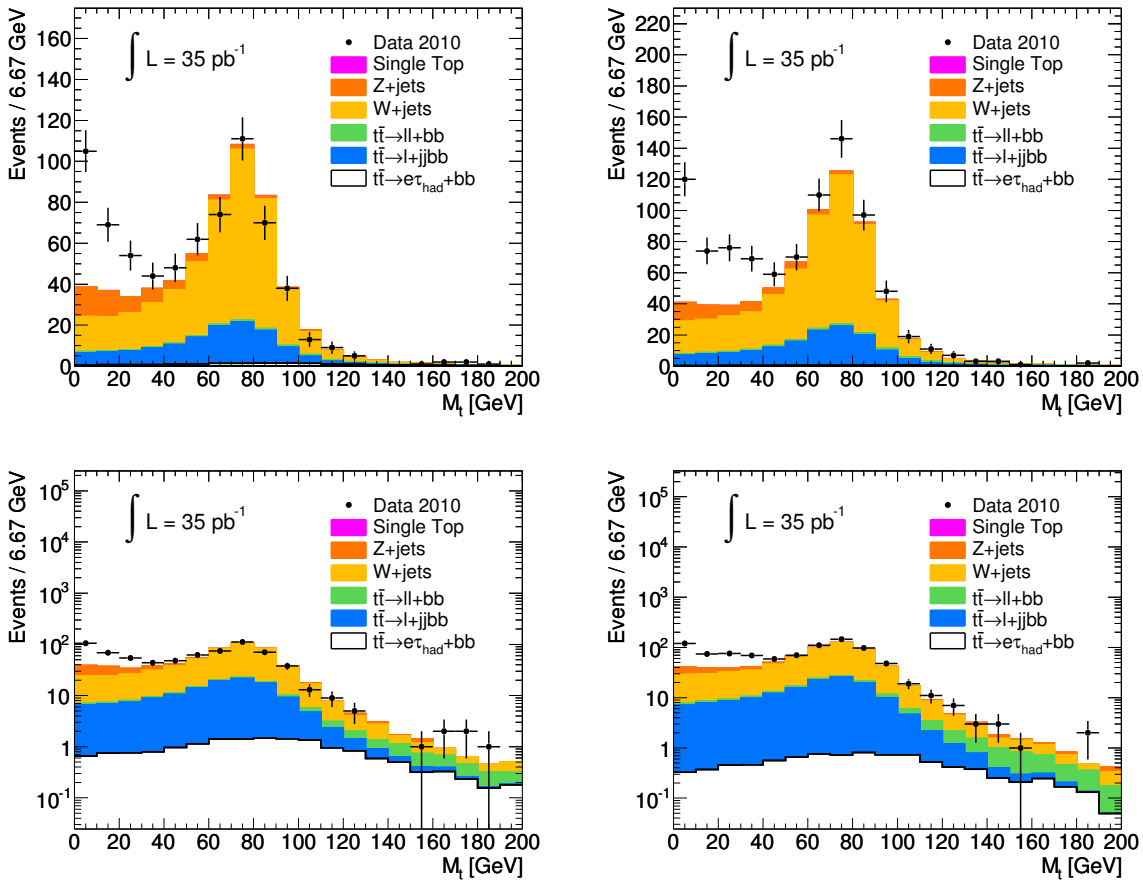


Figure 4.2: *The transverse mass distributions of lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ before requiring a b-jet for τ_1 candidates (left) and τ_3 candidates (right) in the $t\bar{t} \rightarrow e + \tau$ channel. The predicted contribution from signal events, low when compared with the background, is shown more clearly in the logaritmic scale plots (bottom). The excess in the data distribution at low M_t is mainly due to the multi-jet background, not included in the MC samples.*

		R_{EM}	R_{trk}	$1/f_{trk,1}$
τ_1	$p_{\mathrm{T}} < 80$ GeV	$0.0522 - 1.71e^{-4} \cdot p_{\mathrm{T}} + 0.715/p_{\mathrm{T}}$	$0.134 - 2.72e^{-4} \cdot p_{\mathrm{T}} + 0.0328/p_{\mathrm{T}}$	8.33
τ_1	$p_{\mathrm{T}} > 80$ GeV	0.0475	0.113	8.33
τ_3	$p_{\mathrm{T}} < 80$ GeV	$0.0833 - 2.15e^{-4} \cdot p_{\mathrm{T}} + 0.930/p_{\mathrm{T}}$	$0.0146 - 7.21e^{-5} \cdot p_{\mathrm{T}} + 0.879/p_{\mathrm{T}}$	2.5
τ_3	$p_{\mathrm{T}} > 80$ GeV	0.0777	0.0198	2.5

Table 4.3: *Criteria for Safe Cuts τ identification.*

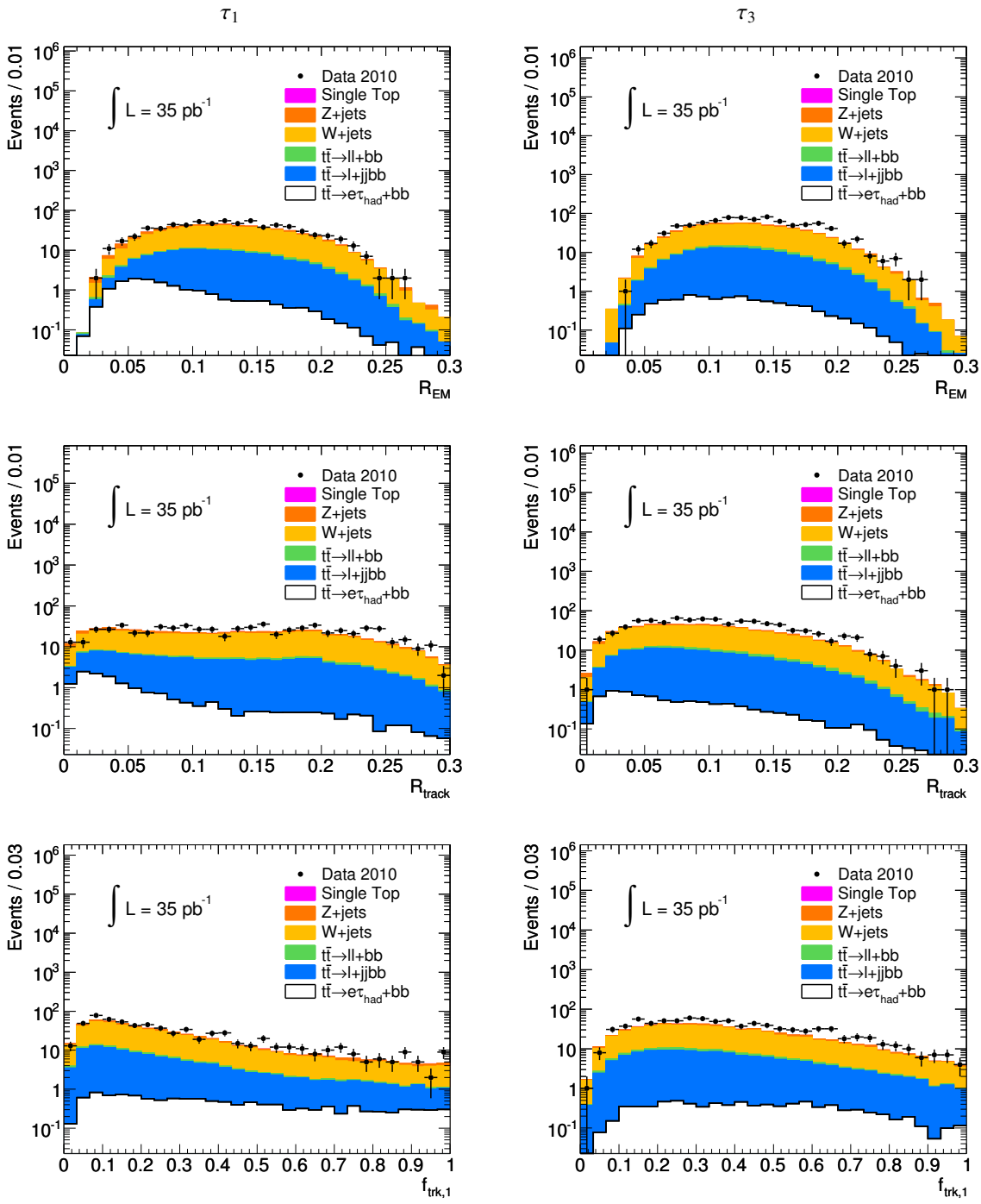


Figure 4.3: *Tau Safe Cuts variables after event selection before requiring a b-jet for τ_1 candidates (left column) and τ_3 candidates (right column) in the $t\bar{t} \to e + \tau$ channel. The excess in the data distribution is mainly due to the multi-jet background, not included in the MC samples.*

τ_1 and τ_3 candidates are shown in Figure 4.4 before (top row) and after (bottom row) applying the *b*-tagging requirement.

The excess in the number of events in data compared to the number expected from MC, which appears in the safe variables and BDT_j plots, is mainly due to the QCD multi-jet background, which is not included in the MC samples.

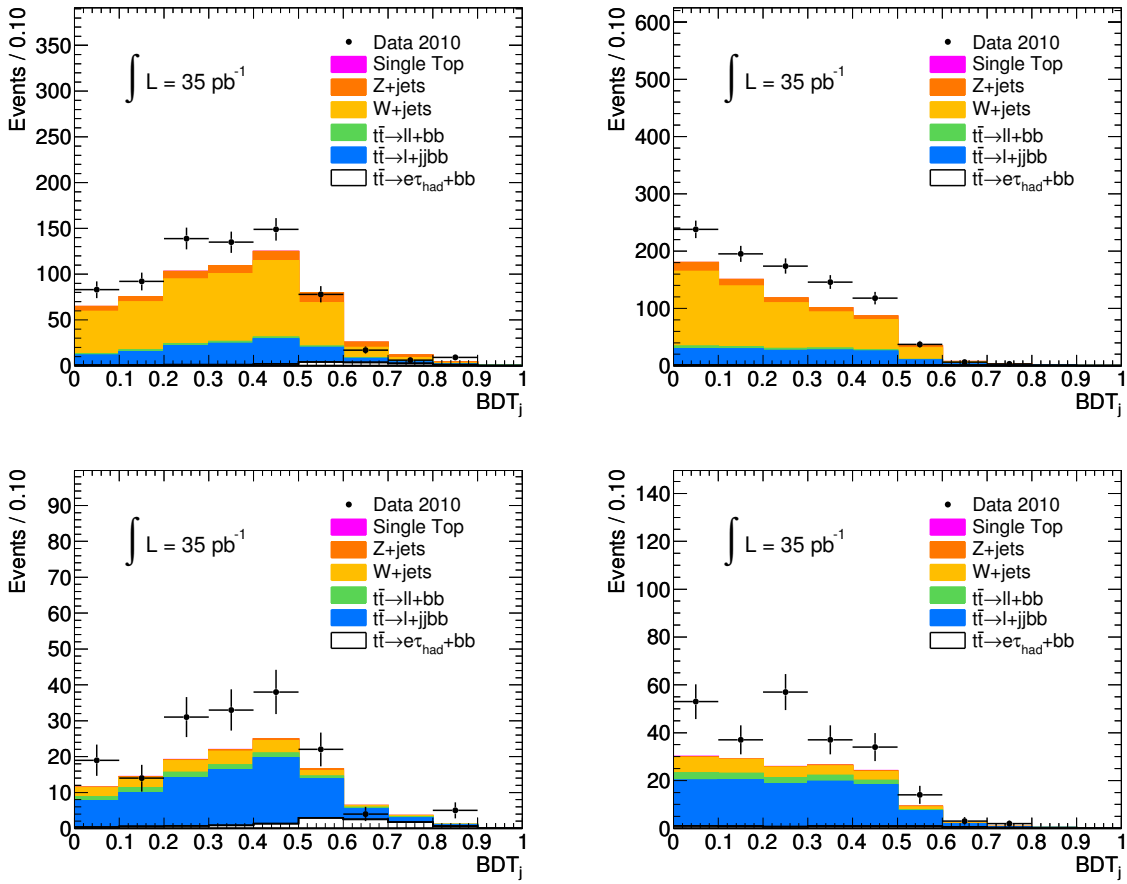


Figure 4.4: *BDT_j distribution after event selection before (top row) and after (bottom row) requiring a b-jet for τ_1 (left) and τ_3 (right) candidates in the $t\bar{t} \rightarrow e + \tau$ channel. The QCD multi-jet background, not included in the MC samples, accounts for the excess in the number of data events.*

For lower values of BDT_j background events predominate. In the region $0.5 < BDT_j < 0.6$ the proportion of signal and background varies rapidly and for BDT_j values greater than 0.6 the signal becomes the major contribution. Lowering the cut from 0.6 to 0.5 increases the background by a factor of three while increasing the signal acceptance only by 30%. Therefore, the tau identification selection cut was set at $BDT_j > 0.6$ for the 2010 data analysis.

The number of MC expected events and the number of observed events in data, after *b*-tagging

and τ identification, using both safe cuts and boosted decision trees, are summarized in Table 4.4 for τ_1 candidates and Table 4.5 for τ_3 candidates. Events are split according to the relative charge between the tau candidate and the electron. Electrons and taus in $t\bar{t} \to \ell + \tau$ signal events have opposite sign charge (OS) while same sign (SS) events are all background.

	$t\bar{t}(\ell,\tau)$	$t\bar{t}(\ell$+jets)	$t\bar{t}(\ell\ell')$	Z+jets	W+jets	$t(\to Wb)q$	Total	Data
Opposite Sign (OS)								
≥ 1 bjet	10.25 ± 0.23	48.85 ± 0.51	4.77 ± 0.16	1.24 ± 0.33	10.06 ± 0.48	0.22 ± 0.01	75.38 ± 0.83	97
$\geq 1\ \tau$ (SC)	4.23 ± 0.15	3.34 ± 0.13	0.69 ± 0.06	0.09 ± 0.09	0.55 ± 0.11	0.01 ± 0.00	8.92 ± 0.25	11
$\geq 1\ \tau$ (BDT)	4.94 ± 0.16	3.31 ± 0.13	0.76 ± 0.06	0.18 ± 0.13	0.46 ± 0.10	0.01 ± 0.00	9.66 ± 0.27	9
Same Sign (SS)								
≥ 1 bjet	1.54 ± 0.09	29.96 ± 0.40	3.42 ± 0.14	1.23 ± 0.33	8.03 ± 0.43	0.17 ± 0.01	44.35 ± 0.69	69
$\geq 1\ \tau$ (SC)	0.08 ± 0.02	1.33 ± 0.08	0.12 ± 0.02	0.00 ± 0.00	0.34 ± 0.09	0.01 ± 0.00	1.88 ± 0.12	2
$\geq 1\ \tau$ (BDT)	0.07 ± 0.02	0.96 ± 0.07	0.11 ± 0.02	0.00 ± 0.00	0.34 ± 0.09	0.00 ± 0.00	1.49 ± 0.12	0

Table 4.4: *Number of opposite sign and same sign events in data and MC before and after applying the Safe Cuts or the $BDT_j > 0.6$ cut for τ_1 identification.*

	$t\bar{t}(\ell,\tau)$	$t\bar{t}(\ell$+jets)	$t\bar{t}(\ell\ell')$	Z+jets	W+jets	$t(\to Wb)q$	Total	Data
Opposite Sign (OS)								
≥ 1 bjet	4.97 ± 0.16	59.89 ± 0.57	6.60 ± 0.19	1.13 ± 0.31	13.59 ± 0.55	0.35 ± 0.01	86.53 ± 0.88	139
$\geq 1\ \tau$ (SC)	1.00 ± 0.07	2.17 ± 0.11	0.11 ± 0.03	0.00 ± 0.00	0.46 ± 0.10	0.01 ± 0.00	3.76 ± 0.17	4
$\geq 1\ \tau$ (BDT)	1.24 ± 0.08	1.03 ± 0.08	0.08 ± 0.02	0.00 ± 0.00	0.23 ± 0.07	0.01 ± 0.00	2.59 ± 0.13	4
Same Sign (SS)								
≥ 1 bjet	1.97 ± 0.10	39.79 ± 0.46	6.55 ± 0.19	1.05 ± 0.30	11.95 ± 0.52	0.26 ± 0.01	61.57 ± 0.79	98
$\geq 1\ \tau$ (SC)	0.05 ± 0.01	0.77 ± 0.06	0.08 ± 0.02	0.18 ± 0.13	0.34 ± 0.09	0.00 ± 0.00	1.43 ± 0.17	3
$\geq 1\ \tau$ (BDT)	0.04 ± 0.01	0.46 ± 0.05	0.05 ± 0.02	0.09 ± 0.09	0.13 ± 0.05	0.00 ± 0.00	0.77 ± 0.12	1

Table 4.5: *Number of opposite sign and same sign events in data and MC before and after applying the Safe Cuts or the $BDT_j > 0.6$ cut for τ_3 identification.*

As we can see, after applying all event selections, b-tagging and τ identification, nearly half of the events are expected to be signal events, while most of the other half of background events come from $t\bar{t}$ events where a jet or a electron are misidentified as a tau. The multi-jet background has been reduced to one event or less after τ identification.

To measure the number of signal events in our data sample we quantify the background component after applying τ identification. The background due to electrons faking a tau is estimated from MC, and the background from jets misidentified as τ candidates (including the contribution from multi-jet events) is estimated using a data driven matrix method technique, described in Section 4.3.

4.3 Background estimation with a Matrix Method

We can apply a simple matrix method to estimate the signal and background components in the data sample [131]. The number of events in the data sample after applying the event selections described in 4.1 and b-tagging is

$$N_{\text{data}}^{\text{loose}} = N_{\text{real}}^{\text{loose}} + N_{\text{fake}}^{\text{loose}}, \tag{4.2}$$

where the "real" subscript indicates the contribution from real taus and "fake" the contribution from jets misidentified as τ candidates. After applying τ identification the composition of the data sample is

$$N_{\text{data}}^{\text{tight}} = N_{\text{real}}^{\text{tight}} + N_{\text{fake}}^{\text{tight}}. \tag{4.3}$$

We define the τ fake rate as the probability that a false τ candidate passes the tau identification criteria: $\varepsilon_{\text{fake}}$. The probability that a real τ passes the tau identification is labelled $\varepsilon_{\text{real}}$.

$$\varepsilon_{\text{real}} = \frac{N_{\text{real}}^{\text{tight}}}{N_{\text{real}}^{\text{loose}}}, \qquad \varepsilon_{\text{fake}} = \frac{N_{\text{fake}}^{\text{tight}}}{N_{\text{fake}}^{\text{loose}}}. \tag{4.4}$$

Knowing the tau identification efficiencies, we can solve the system of equations 4.2-4.4 to determine the background prediction for events with a jet misidentified as a τ:

$$N_{\text{fake}}^{\text{tight}} = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}}(N_{\text{data}}^{\text{loose}} * \epsilon_{\text{real}} - N_{\text{data}}^{\text{tight}}). \tag{4.5}$$

$\varepsilon_{\text{real}}$ is estimated from $t\bar{t} \to \ell + \tau$ MC events with a real tau and $\varepsilon_{\text{fake}}$ is derived from data using a γ+ jet selection as described in section 4.3.1.

4.3.1 Tau fake rates using γ+ jets events

Events with a jet misidentified as a τ lepton ($W \to \ell\nu$, $t\bar{t} \to \ell$ + jets) are the dominant background for $t\bar{t} \to \ell + \tau$. According to MC simulation, the jet composition of τ fakes in the $t\bar{t}$ sample after all event selections and b-tagging are applied, shown in Table 4.6, is mainly formed of light-quark jets, but gluon and b-jets are also present and non negligible.

Jet Composition of Jet-to-τ Fake Candidates

τ_1	gluon	quark	b	τ	e	μ
$t\bar{t}$ btag (SS)	7%	77%	12%	0%	3%	0%
$t\bar{t}$ btag (OS)	5%	77%	7%	9%	1%	1%
τ_3	gluon	quark	b	τ	e	μ
$t\bar{t}$ btag (SS)	7%	72%	19%	0%	1%	0%
$t\bar{t}$ btag (OS)	5%	77%	14%	2%	1%	0%

Table 4.6: *Composition of all τ candidates in MC events, after requiring the selections detailed in the text. OS (SS) stands for the opposite (same) charge sign between the electron and τ candidate.*

Light-quark jets have a higher probability of faking a τ than all other jet types because of their narrow shower width and low multiplicity. Therefore, the τ fake rate depends on the type of jet and it is essential to understand its behaviour using real data. The determination of the fake τ jet composition in the 35 pb^{-1} 2010 data sample is complicated due to the limited statistics available. Therefore, the approach taken with the early ATLAS data was to set an upper limit on the background from fake τ leptons using samples rich in light-quark jets as they have a higher probability of faking a τ than other types of jet.

A control sample of γ + jets from data provides a light quark dominated sample without any appreciable contamination from real τ leptons with which the τ fake rate can be estimated.

Therefore, the fake rate measured from γ + jets consists mostly of fake τ leptons initiated by light-quark jets, and relying on γ + jets to estimate the jet contamination in a sample of τ candidates leads to an overestimate of the background.

To obtain a γ + jets sample from data we apply the following selections:

- events have to be included in the GRL and pass event cleaning criteria (at least one vertex with 4 associated tracks)
- events are required to pass a single γ trigger.
- At least one identified photon with $p_{\mathrm{T}}^{\gamma} > 15$ GeV and $|\eta| < 2.5$ (excluding the crack region $1.37 < \eta < 1.52$). Identified photons must pass a tight photon selection and be well isolated [132]. The total energy in a cone of radius 0.4 around the γ is less than 3 GeV.
- At least one good jet in the event with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$. The jet must overlap with a reconstructed τ candidate.
- The photon and the τ candidate must be separated by $\Delta\phi > |\pi - 0.2|$ and $\Delta R \geq 0.2$.
- If there is a second jet present in the event, its p_{T} must not be greater than 20% of p_{T}^{γ}.

This selection provides events with a high purity of light-quark initiated tau jet candidates with which we measure $\varepsilon_{\mathrm{fake}}$. The distributions of $\varepsilon_{\mathrm{real}}$ and $\varepsilon_{\mathrm{fake}}$ as a function of the tau p_{T} are shown in figure 4.5 for τ_1 (left) and τ_3 (right) candidates using safe cuts (top) and BDT_j (bottom) based identification.

4.3.2 Matrix Method results

The calculated $\varepsilon_{\mathrm{real}}$ and $\varepsilon_{\mathrm{fake}}$ are applied to the number of tau candidates in the b-tagged data samples to estimate the jet background using equation 4.5.

The number of observed events and background predictions using the matrix method are shown in Table 4.7 for τ_1 candidates and Table 4.8 for τ_3 candidates. The prediction of jet fakes from data is compared to the expected background from MC.

	Observed Events	Expected background	Measured background
		τ_1 Opposite Sign	
SC	11	4.6 ± 0.2	5.9 ± 1.1
BDT	9	4.5 ± 0.2	5.4 ± 0.9
		τ_1 Same Sign	
SC	2	1.7 ± 0.1	5.3 ± 0.9
BDT	0	1.4 ± 0.1	4.7 ± 0.7

Table 4.7: *Background estimate for $t\bar{t} \rightarrow e + \tau_1$ using a Matrix method. Safe Cuts and BDT tau identification results are shown and compared to MC prediction.*

Since $N_{\mathrm{fake}}^{\mathrm{tight}}$ is obtained using fake rates from a γ + jets light-quark sample, which is expected to have a higher fake rate than any other mixture of jets, what we obtain is actually an upper limit for the background in our sample. The OS sample measurements are closer to expectations

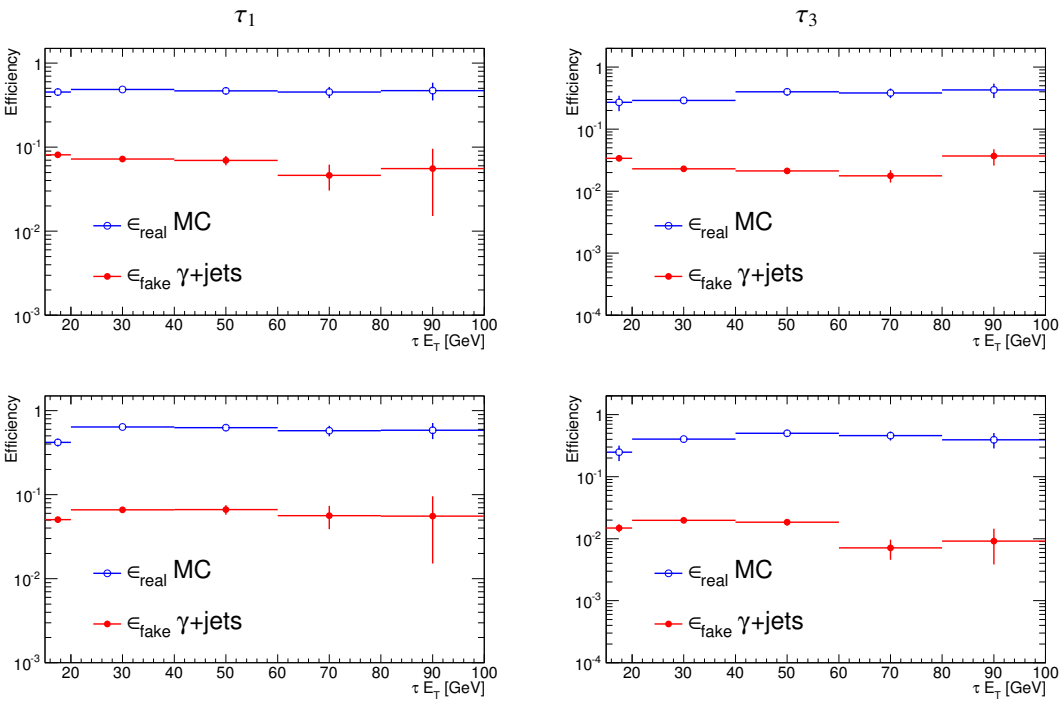


Figure 4.5: *Tau identification efficiency for real tau candidates in MC and fake tau candidates form γ + jets events in data. Safe cuts efficiencies are plotted in the top row, BDT_j efficiencies are plotted on the bottom row.*

	Observed Events	Expected background	Measured background
		τ_3 Opposite Sign	
SC	3	2.3 ± 0.1	3.3 ± 0.5
BDT	4	1.3 ± 0.1	2.2 ± 0.3
		τ_3 Same Sign	
SC	2	1.1 ± 0.1	2.2 ± 0.4
BDT	1	0.7 ± 0.1	1.5 ± 0.4

Table 4.8: *Background estimate for $t\bar{t} \rightarrow e + \tau_3$ using a Matrix method. Safe Cuts and BDT tau identification results are shown and compared to MC prediction.*

than the measurement for the SS sample. This was expected since the light-quark contribution to the jet background is higher in the OS sample and SS events have a significantly larger gluon content.

The background is, then, significantly overestimated and it can not be used to measure the $t\bar{t}$ cross section in the $\ell + \tau$ channel. Alternative methods to estimate the background are derived

using the 2011 dataset and will be detailed in chapter 5.

4.4 Summary and conclusions

Two different τ identification criteria have been studied to test the sensitivity of τ identification techniques in $t\bar{t}$ events. All the data used correspond to pp collisions with a center-of-mass energy of $\sqrt{s} = 7$ TeV taken between March 30th and October 24th 2010. The total integrated luminosity of the sample is 35.3 pb^{-1}.

Cut-based tau identification provides a more conservative approach to τ-jet separation. BDT tau identification is designed to maximize signal and background separation by recycling events that both pass and fail cuts. The similarity of the jet background to the tau signal is a motivation to study the use of boosted decision trees, as they combine many weak classifying variables into a stronger discriminant for tau identification in ATLAS. This method will be used in the analysis of the 2011 data, described in chapter 5.

5

Measurement of the $t\bar{t}$ cross section in the $\ell + \tau$ channel using the Fitting Method

5.1 Methodology of the analysis

The $t\bar{t}$ cross section ($\sigma_{t\bar{t}}$) is proportional to the probability that a pair of top quarks are produced in a pp collision at the LHC. This cross section is related to the number of $t\bar{t}$ events produced ($N_{t\bar{t}}$) as

$$\sigma_{t\bar{t}} = \frac{N_{t\bar{t}}}{\mathcal{L}}, \tag{5.1}$$

where $\mathcal{L}$ is the integrated luminosity. Only a small fraction (BR) of these events, about a 5%, contain a lepton (electron or muon) and a tau in their final state. By applying an event selection criteria that aims simultaneously to enrich the signal content and maximize the background rejection, we obtain a number of candidate signal events that can be expressed as

$$N_{\text{signal}} = N_{t\bar{t}} \cdot BR \cdot A \cdot \epsilon = N_{\text{data}} - N_{\text{bkg}}, \tag{5.2}$$

where A accounts for the geometrical acceptance of the ATLAS detector to signal events and ϵ includes the trigger and object reconstruction efficiency and also the efficiency of the event selection criteria. N_{data} is the number of events which passes the event selection in a data sample of integrated luminosity $\mathcal{L}$ and receives contributions from all the Standard Model processes, signal and background, with a similar experimental signature in the ATLAS detector.

Therefore, the $t\bar{t} \to \ell + \tau$ cross section is calculated as

$$\sigma_{t\bar{t}\to\ell+\tau} = \frac{N_{\text{signal}}}{A \cdot \epsilon} \cdot \frac{1}{\mathcal{L}}. \tag{5.3}$$

Signal events come from top quark pair production in the lepton plus tau final state:

$$t\bar{t} \rightarrow W(\rightarrow e/\mu + \nu_{e/\mu})bW(\rightarrow \tau + \nu_{\tau})b. \tag{5.4}$$

The background processes affecting the $t\bar{t}(\ell, \tau)$ signal are:

- Top quark pair production in the final states:
  1. lepton plus jets: $t\bar{t} \rightarrow W(\rightarrow e/\mu + \nu_{e/\mu})bW(\rightarrow qq^{'})b$
  2. dilepton: $t\bar{t} \rightarrow W(\rightarrow e/\mu + \nu_{e/\mu})bW(\rightarrow e/\mu + \nu_{e/\mu})b$
- Single top quark production
- W and Z boson production with jets associated
- Diboson pair production WW, WZ, ZZ

We have to differentiate these background processes (N_{bkg}) to be able to select the signal events (N_{signal}) and measure the cross section. Therefore, to measure the $t\bar{t}$ cross section in the $\ell + \tau$ channel two key steps are taken:

1. We apply an object and event selection to extract $t\bar{t}$ events from data, optimized to maximize the number of signal events against the background.
2. We estimate the background with the lowest possible uncertainty.

The kinematic selection criteria applied for the analysis (see Section 5.2) will favour events with one W decaying to a charged lepton (electron or muon), either directly from the W or from a leptonic τ decay, and a neutrino and the other W decaying to a τ lepton successively decaying hadronically. At least one of the two b-quark jets is identified using an algorithm that can identify b-jets with high efficiency while maintaining a high rejection of light-quark jets.

After applying all the kinematic selections and b-tagging identification, the dominant background is the top quark pair production in the lepton plus jets channel in which the τ candidate comes from jets misidentified as hadronic τ decays. Therefore, τ lepton identification is the critical discriminator between signal and background. In this analysis we will use the τ lepton identification technique based on Boosted Decision Trees, described in Section 3.3.3.

The composition of the data sample, after applying the proper object selection on leptons, jets, b-jets, $E_{\text{T}}^{\text{miss}}$ and tau followed by a customized event selection to enhance the $t\bar{t}(\ell, \tau)$ signal content, can be expressed in terms of the source that produces the τ jet candidate. In MC, this composition can be investigated by analysing the truth object (real τ leptons, b-quark and light-quark jets and gluons) that is matched[1] to the τ candidate (see Figure 5.3).

At this point, if the data sample is divided according to the relative charge of the lepton (electron or muon) and the τ candidate, the contributions to the BDT_j shape can be expresed as follows:

$$\text{OS}_{\text{Data}} = \tau \text{ leptons} + \text{OS}_{\text{q}} + b\text{-quarks} + \text{Gluons}, \tag{5.5}$$

[1] Truth matching relates, in MC, reconstructed particles in the detector with their corresponding particles at generator level (truth particles) by measuring the distance between the reconstructed and the truth particles, $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, and imposing $\Delta R < 0.4$

$$\mathrm{SS_{Data}} = \mathrm{SS_q} + b\text{-quarks} + \mathrm{Gluons}. \tag{5.6}$$

Only events with oppositely charged leptons have contribution from real taus. The gluon and b-quark contributions are charge symmetric, therefore their contribution to the OS and SS data is identical. $\mathrm{OS_q}$ and $\mathrm{SS_q}$ are the light quark component of the OS and SS samples, respectively.

The critical aspect of extracting the signal component of the data sample using BDT_j distributions is that the BDT_j shape depends on the kind of jet. Due to their narrow shower width and lower track multiplicity, light-quark jets have a higher probability of faking a τ lepton than other jet types. Therefore, building a background model which properly reflects the jet composition is crucial in order to correctly estimate the fake τ contamination in the signal region.

A first approach to this analysis, done with the first 1.08 fb^{-1} [133], [134], and 2.05 fb^{-1} [135], [67] of ATLAS data, uses the OS-SS technique to get a data sample where our signal is intact but the background is reduced to only light quark jets by subtracting the BDT_j distribution built with the SS data from the BDT_j distribution built with the OS data. Then the BDT_j distribution of the τ candidates in the data sample is fitted to a signal and a light-quark background template [136]. A cross check measurement is made applying a cut on the BDT_j discriminant and using a matrix method to normalize the background, following the method developed in [137]. Both the OS-SS technique and the cross check measurement with the Matrix Method are described in detail in Appendix A.

A potential problem of the OS-SS technique is that the proportion of light quark jets falsifying the τ candidates with opposite reconstructed charge with respect to the original charge of the light quark that initiated the jet, can not be determined in the control samples used to derived the templates. Moreover, these cases will migrate to the SS category and vice versa. Furthermore, since the fragmentation will be different from jets contributing to OS from jets contributing to SS, the shape of the BDT_j distribution built from them will differ. On top of that, the mixture of OS and SS light-quark jets also does not necessarily have the same proportion in the sample used to derive the background template and the sample from which the signal is extracted.

Another drawback of the OS-SS technique is that, for a given amount of data or integrated luminosity, the subtraction of BDT_j distributions results in an increased statistical error.

A different approach to the fitting technique is presented in this thesis aiming to improve the existing background model of jets faking the taus affecting the top pair lepton plus tau channel. We build templates for the gluon and light-quark jet BDT_j distributions and then fit the OS data BDT_j distribution to this templates and a signal BDT_j template obtained from the MC samples:

$$\mathrm{OS_{Data}} = k_s \cdot \mathrm{signal} + k_g \cdot \mathrm{Gluons} + k_q \cdot \mathrm{OS_q}. \tag{5.7}$$

This way we reduce the statistical error of the cross section measurement since we are not subtracting the SS events from the OS sample and have a better modelling for the $t\bar{t} \to \ell + \tau$ background.

The fitting technique is described in Section 5.3. The derivation of the background templates from data is detailed in section 5.4. The fitting procedure is elaborated in Section 5.5. It is first applied to a MC test sample to test its robustness and afterwards applied to the data sample to measure the number of signal events. Sources of systematic errors are described and taken into account for the final cross section measurement in Sections 5.6 and 5.7, respectively.

5.2 Event selection

The event selection for 2011 analysis has been optimized to try to isolate the event topology of our signal events: high E_{T} b-jets, 1 high p_{T} lepton, 1 hadronically decaying τ and large $E_{\mathrm{T}}^{\mathrm{miss}}$, and simultaneously to reduce the backgrounds. The processes that contribute to our $t\bar{t} \to \ell + \tau$ signal events include $t\bar{t} \to W(\to \tau_{had})bW(\to e/\mu)b$ and τ leptonic decays: $t\bar{t} \to W(\to \tau_{had})bW(\to \tau \to e/\mu)b$. The main backgrounds correspond to events with jets falsifying the tau signal (W + jets, semileptonic $t\bar{t}$ and QCD multi-jet events) and non $t\bar{t}$ events containing real τ leptons ($Z \to \tau\tau$, single top, mainly the $Wt(\to Wb)$ production and WW, WZ, ZZ diboson production).

We start the event selection process for $t\bar{t} \to \ell + \tau$ events requiring events selected online by a single lepton trigger: a single-muon trigger with a p_{T} threshold of 18 GeV or a single-electron trigger with a p_{T} threshold of 20 GeV, rising to 22 GeV during periods of high instantaneous luminosity. These triggers are similar to the offline object reconstruction, but with looser selection criteria, and reach their respective efficiency plateaus at 25 GeV (electrons) and 20 GeV (muons).

The offline preselection is based on data quality criteria and optimized using Monte Carlo samples:

- Data cleaning is needed in order to ensure the quality of object reconstruction. Events not included in the Good Run List are discarded from the analysis.

- In order to ensure the event originated from the collision and to reject the non-collision background events, an event is required to have a primary vertex with at least five tracks, each one with an associated $p_{\mathrm{T}} > 400$ MeV, and it is discarded if any jet with $p_{\mathrm{T}} > 20$ GeV fails the jet quality selections designed to reject jets arising from out-of-time activity or calorimeter noise [130].

- There must be one, and only one, isolated muon (electron) and no identified electrons (muons). This requirement greatly reduces the multi-jet background.

- There must be at least one loose τ candidate, as defined in Section 3.2.

- There must be at least two jets with $p_{\mathrm{T}} > 25$ GeV not overlapping with a τ candidate. The overlapping jet is the closest jet to the τ candidate within $\Delta R(\tau, jet) < 0.4$. In the case of two τ candidates, at least one jet is required to be not overlapping with either τ candidate, and both candidates are kept until a tight selection is made on the τ candidate.

- A cut on the transverse missing energy, $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV, is applied to reduce the QCD multi-jet background. Due to the large cross section for multi-jet production, the background from jets misidentified as isolated electron or muons is not negligible, but we can reduce it by requiring a significant missing transverse momentum to account for the presence of energetic neutrinos in our signal.

- A cut on the scalar sum of the lepton p_{T}, the p_{T} of the jets above 25 GeV, the p_{T} of the τ candidate, and $E_{\mathrm{T}}^{\mathrm{miss}}$, $H_{\mathrm{T}} > 200$ GeV, is applied to reduce the W + jets background.

- There must be at least one jet identified as a b-jet using the CombNN algorithm at a 70% efficiency point, cutting at a value of 0.35 in the corresponding jet weight.

This preselection defines what we will be calling the ≥ 1 b-tag signal region.

In the case of multiple τ candidates, all of them are used for both signal and background templates in the fitting method. Cut flow Tables 5.1-5.4 and control plots 5.1 and 5.2 are filled per event, considering only the so called "leading τ", which is the τ candidate most likely to pass the tau identification. In case of having both τ_1 and τ_3 candidates, the leading τ is the τ_1. If all τ candidates are 1-prong or 3-prong, then the leading τ is the one with the highest BDT_j value.

The number of expected and observed events after each selection cut are shown in Tables 5.1 and 5.2 for the $\mu + \tau$ channel and in Tables 5.3 and 5.4 for the $e + \tau$ channel. After all pre-selections have been done, 70% of events come from $t\bar{t}$ processes, 20% from W + jets and approximately 6% from single top processes. Therefore, the dominant backgrounds are W + jet and $t\bar{t} \to \ell$ + jets, where a jet is misidentified as a τ lepton. Applying the b-tagging requirement suppresses the W + jet background, which leaves $t\bar{t} \to \ell$ + jets as the dominant background. At this stage only τ lepton identification allow us to separate signal form background in our data sample.

Table 5.1: *$\mu + \tau$ cut flow for τ_1 candidates. $t\bar{t}(\ell\ell')$ are $t\bar{t}$ dilepton events with only one lepton reconstructed as a muon and a lepton or jet reconstructed as a τ. The last two rows show the numbers individually for opposite signed (OS) and same signed (SS) events which pass the final b-jet selection. The errors are the statistical uncertainties.*

Cut	$t\bar{t}(\mu,\tau)$	$t\bar{t}(\ell + \text{jets})$	$t\bar{t}(\ell\ell')$	Z+jets	W+jets	Single top	Diboson	Total	Data
Isolated μ	3972 ± 14	25938 ± 37	3621 ± 14	739203 ± 698	8398984 ± 9931	11663 ± 57	12042 ± 60	9195423 ± 9956	15339738
≥ 1 τ candidate	784 ± 6	3346 ± 13	169 ± 3	20537 ± 109	61832 ± 395	721 ± 13	1198 ± 19	88586 ± 411	258721
$N_{\text{jet}} \geq 2$	641 ± 6	3189 ± 13	138 ± 3	2193 ± 35	12396 ± 112	379 ± 8	287 ± 9	19224 ± 119	28255
$E_T^{\text{miss}} > 30$ GeV	559 ± 5	2561 ± 12	124 ± 2	958 ± 23	8929 ± 92	303 ± 7	203 ± 8	13637 ± 96	15700
$H_T + E_{\text{T}}^{\text{miss}} > 200$ GeV	550 ± 5	2546 ± 12	123 ± 2	764 ± 21	7614 ± 80	290 ± 7	177 ± 7	12065 ± 85	12997
≥ 1 bjet	472 ± 5	2105 ± 11	98 ± 2	51 ± 5	712 ± 31	204 ± 6	16 ± 2	3658 ± 35	3809
≥ 1 bjet (OS)	431 ± 5	1506 ± 9	54 ± 2	33 ± 4	440 ± 24	147 ± 5	10 ± 2	2621 ± 27	2475
≥ 1 bjet (SS)	41 ± 1	599 ± 6	44 ± 1	17 ± 3	272 ± 21	57 ± 3	6 ± 1	1037 ± 22	1334

Table 5.2: *$\mu + \tau$ cut flow for τ_3 candidates. $t\bar{t}(\ell\ell')$ are $t\bar{t}$ dilepton events with only one lepton reconstructed as a muon and a lepton or jet reconstructed as a τ. The last two rows show the numbers individually for opposite signed (OS) and same signed (SS) events which pass the final b-jet selection. The errors are the statistical uncertainties.*

Cut	$t\bar{t}(\mu,\tau)$	$t\bar{t}(\ell + \text{jets})$	$t\bar{t}(\ell\ell')$	Z+jets	W+jets	Single top	Diboson	Total	Data
Isolated μ	3972 ± 14	25938 ± 37	3621 ± 14	739203 ± 698	8398984 ± 9931	11663 ± 57	12042 ± 60	9195423 ± 9956	15339738
≥ 1 τ candidate	662 ± 6	8210 ± 21	522 ± 5	29086 ± 132	156968 ± 588	1867 ± 21	2246 ± 26	199562 ± 604	700853
$N_{\text{jet}} \geq 2$	508 ± 5	7733 ± 20	413 ± 5	3820 ± 45	31896 ± 178	892 ± 13	569 ± 13	45831 ± 186	71154
$E_T^{\text{miss}} > 30$ GeV	446 ± 5	6203 ± 18	371 ± 4	1525 ± 29	23116 ± 148	705 ± 11	398 ± 11	32765 ± 153	39118
$H_T + E_{\text{T}}^{\text{miss}} > 200$ GeV	440 ± 5	6168 ± 18	366 ± 4	1278 ± 26	20192 ± 131	680 ± 11	347 ± 10	29473 ± 136	33007
≥ 1 bjet	370 ± 4	5141 ± 16	295 ± 4	98 ± 7	1973 ± 50	478 ± 9	42 ± 3	8397 ± 54	9410
≥ 1 bjet (OS)	240 ± 4	3350 ± 13	147 ± 3	46 ± 5	1120 ± 37	309 ± 7	28 ± 3	5240 ± 40	5712
≥ 1 bjet (SS)	129 ± 3	1791 ± 10	149 ± 3	51 ± 5	854 ± 34	169 ± 6	14 ± 2	3157 ± 37	3698

Table 5.3: *$e+\tau$ cut flow for τ_1 candidates. $t\bar{t}(\ell\ell')$ are $t\bar{t}$ dilepton events with only one lepton reconstructed as an electron and a lepton or jet reconstructed as a τ. The last two rows show the numbers individually for opposite signed (OS) and same signed (SS) events which pass the final b-jet selection. The errors are the statistical uncertainties.*

Cut	$t\bar{t}(\mu,\tau)$	$t\bar{t}(\ell+\text{jets})$	$t\bar{t}(\ell\ell')$	Z+jets	W+jets	Single top	Diboson	Total	Data
Isolated e	3604 ± 14	23501 ± 35	3230 ± 13	770763 ± 709	6878437 ± 9081	9792 ± 51	9907 ± 54	7699235 ± 9109	10960533
≥ 1 τ candidate	695 ± 6	3018 ± 13	151 ± 3	21889 ± 114	93486 ± 835	623 ± 11	944 ± 17	120806 ± 843	228810
$N_{\text{jet}} \geq 2$	572 ± 5	2874 ± 12	124 ± 2	3952 ± 47	10604 ± 122	342 ± 8	268 ± 8	18737 ± 132	29306
$E_T^{\text{miss}} > 30$ GeV	495 ± 5	2249 ± 11	111 ± 2	1365 ± 28	7198 ± 100	254 ± 7	165 ± 7	11837 ± 105	14033
$H_T + E_{\text{T}}^{\text{miss}} > 200$ GeV	490 ± 5	2238 ± 11	110 ± 2	1199 ± 26	6128 ± 77	246 ± 6	147 ± 6	10558 ± 83	11926
≥ 1 bjet	416 ± 5	1850 ± 10	89 ± 2	64 ± 6	540 ± 26	171 ± 5	14 ± 2	3143 ± 29	3373
≥ 1 bjet (OS)	377 ± 4	1307 ± 8	51 ± 2	46 ± 5	316 ± 21	122 ± 4	9 ± 1	2227 ± 24	2269
≥ 1 bjet (SS)	40 ± 1	543 ± 5	38 ± 1	17 ± 3	224 ± 15	49 ± 3	5 ± 1	916 ± 17	1104

Table 5.4: *$e+\tau$ cut flow for τ_3 candidates. $t\bar{t}(\ell\ell')$ are $t\bar{t}$ dilepton events with only one lepton reconstructed as an electron and a lepton or jet reconstructed as a τ. The last two rows show the numbers individually for opposite signed (OS) and same signed (SS) events which pass the final b-jet selection. The errors are the statistical uncertainties.*

Cut	$t\bar{t}(\mu,\tau)$	$t\bar{t}(\ell+\text{jets})$	$t\bar{t}(\ell\ell')$	Z+jets	W+jets	Single top	Diboson	Total	Data
Isolated e	3604 ± 14	23501 ± 35	3230 ± 13	770763 ± 709	6878437 ± 9081	9792 ± 51	9907 ± 54	7699235 ± 9109	10960533
≥ 1 τ candidate	589 ± 6	7437 ± 20	471 ± 5	32595 ± 140	237909 ± 1311	1597 ± 19	1928 ± 24	282526 ± 1319	630596
$N_{\text{jet}} \geq 2$	453 ± 5	7015 ± 19	371 ± 4	8261 ± 69	28143 ± 217	836 ± 12	536 ± 12	45613 ± 229	73435
$E_T^{\text{miss}} > 30$ GeV	395 ± 5	5502 ± 17	331 ± 4	2612 ± 38	19262 ± 175	635 ± 11	327 ± 10	29065 ± 180	34471
$H_T + E_{\text{T}}^{\text{miss}} > 200$ GeV	392 ± 5	5476 ± 17	329 ± 4	2409 ± 37	16984 ± 155	619 ± 11	294 ± 9	26502 ± 161	29930
≥ 1 bjet	329 ± 4	4544 ± 16	263 ± 4	148 ± 9	1707 ± 50	435 ± 9	32 ± 3	7458 ± 54	8218
≥ 1 bjet (OS)	214 ± 3	2955 ± 13	132 ± 3	83 ± 7	965 ± 36	272 ± 7	17 ± 2	4639 ± 40	5030
≥ 1 bjet (SS)	115 ± 2	1589 ± 9	131 ± 3	66 ± 6	742 ± 34	163 ± 5	14 ± 2	2820 ± 36	3188

5.2.1 QCD Multi-jet modelling

QCD multi-jet events that contribute to the signal region must have a jet faking the isolated electron or muon and another jet faking the tau candidate.

In cut flow tables 5.1-5.4 the data entries contain more events than the MC expectation, especially in the first selections, prior to the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut. These differences are mainly attributed to the QCD multi-jet background, which is not included in the above MC expectations. The QCD multi-jet background is derived from data given that the detector simulation is insufficiently precise to model the rate of jets falsifying isolated leptons (electrons or muons) and tau candidates.

The shape of the QCD multi-jet background is modelled using a data sample that is enriched in QCD events, obtained with an identical event selection criteria as the one applied to the ≥ 1 b-jet signal region but reversing the isolation criteria of the electron or muon. The contribution of other SM processes to this sample with anti-isolated leptons is modelled using MC and subtracted.

As the acceptance for these non-isolated events is not expected to equal that of multi-jets in the isolated sample, the overall normalization of the events is derived using a fit. The fitting is performed on the data distribution of the transverse mass

$$M_{\mathrm{T}}(E_{\mathrm{T}}^{\mathrm{miss}}, \ell) = \sqrt{(E_{\mathrm{T}}^{\ell} + E_{\mathrm{T}}^{\mathrm{miss}})^2 + (p_x^{\ell} + E_x^{\mathrm{miss}})^2 + (p_y^{\ell} + E_y^{\mathrm{miss}})^2}, \tag{5.8}$$

with the multi-jet template and the rest of the SM MC over the full range from 0 to 200 GeV. The fit floats the normalization of the QCD model and the non-QCD processes (MC) individually to the data using a χ^2 minimalization. The underlying assumption is that the shape of the transverse mass distribution for QCD is the same in the baseline and the inverted selection. The fitting is performed at each of the event selection points of interest ($\mathrm{N_{jet}}$, $E_{\mathrm{T}}^{\mathrm{miss}}$, H_{T}, and ≥ 1 b-tag) and individually for τ_1 and τ_3 and the μ and e channels. Additionally, the fits are performed separately on the distributions in which the τ and lepton have a relative charge of the opposite sign (OS) or the same sign (SS). An uncertainty of 30% is applied to the QCD distributions, which was evaluated in [138].

The resulting multi-jet normalization for each fit is shown in table 5.5. The multi-jet plus MC normalization is also compared to the data and agrees well. These normalization factors have not been explicitly calculated in this thesis work; rather, they have been obtained from section 5.3 of [135].

Figures 5.1 and 5.2 show the τ candidates BDT_j distributions before and after applying b-tagging. The agreement between data and MC is quite good when the QCD multi-jet background in taken into account.

The background models for this analysis will be obtained from data (see section 5.4). Therefore, the MC and QCD multi-jet background distributions and uncertainties do not affect the cross section measurement. The multi-jet distribution has been added to the control plots to ensure we understand the event topology before proceeding any further with the analysis.

Table 5.5: *Number of multi-jet events estimated in data using a fit on the $M_{\mathrm{T}}(E_{\mathrm{T}}^{\mathrm{miss}}, \ell)$ distribution at each cut stage and individually for OS and SS, τ_1 and τ_3, and the μ and e channels. The total background estimation, shown in the 4th column, is given by the addition of the multi-jet (OS+SS) normalization and the total MC from Tables 5.1-5.4. This is compared to the data in the last column. The uncertainty on the multi-jet normalization is 30%.*

$\mu + \tau_1$	OS multi-jet	SS multi-jet	multi-jet+MC	Data
$N_{\mathrm{jet}} \geq 2$	4693	4368	28285 ± 1927	28255
$E_T^{\mathrm{miss}} > 30$ GeV	1119	1056	15812 ± 471	15700
$H_T > 200$ GeV	687	582	13334 ± 283	12997
≥ 1 $b-$jet	117	146	3921 ± 66	3809
$\mu + \tau_3$	OS multi-jet	SS multi-jet	multi-jet+MC	Data
$N_{\mathrm{jet}} \geq 2$	13191	12724	71746 ± 5501	71154
$E_T^{\mathrm{miss}} > 30$ GeV	2810	2702	38277 ± 1179	39118
$H_T > 200$ GeV	1539	1478	32490 ± 654	33007
≥ 1 $b-$jet	464	401	9262 ± 192	9410
$e + \tau_1$	OS multi-jet	SS multi-jet	multi-jet+MC	Data
$N_{\mathrm{jet}} \geq 2$	5701	5594	30032 ± 2400	29306
$E_T^{\mathrm{miss}} > 30$ GeV	1375	1310	14522 ± 579	14003
$H_T > 200$ GeV	901	860	12319 ± 382	11926
≥ 1 $b-$jet	165	135	3443 ± 70	3373
$e + \tau_3$	OS multi-jet	SS multi-jet	multi-jet+MC	Data
$N_{\mathrm{jet}} \geq 2$	17708	17854	81175 ± 7547	73425
$E_T^{\mathrm{miss}} > 30$ GeV	4268	4150	37483 ± 1795	34471
$H_T > 200$ GeV	3011	2861	32374 ± 1256	29930
≥ 1 $b-$jet	690	606	8754 ± 281	8218

Figure 5.1: *BDT_j distributions of all τ_1 (top) and τ_3 (bottom) candidates in Data and MC for the $\mu + \tau$ channel. On the left plots the BDT_j distributions are shown after the events selections specified in Section 5.2 and before applying the b-tagging requirement. At this stage $t\bar{t} \rightarrow \ell$ + jet and W+ jets are the dominant backgrounds. After applying the b-tagging requirement the W+ jets background has been highly suppressed and $t\bar{t} \rightarrow \ell$ + jet becomes the dominant background, as shown in the right hand plots.*

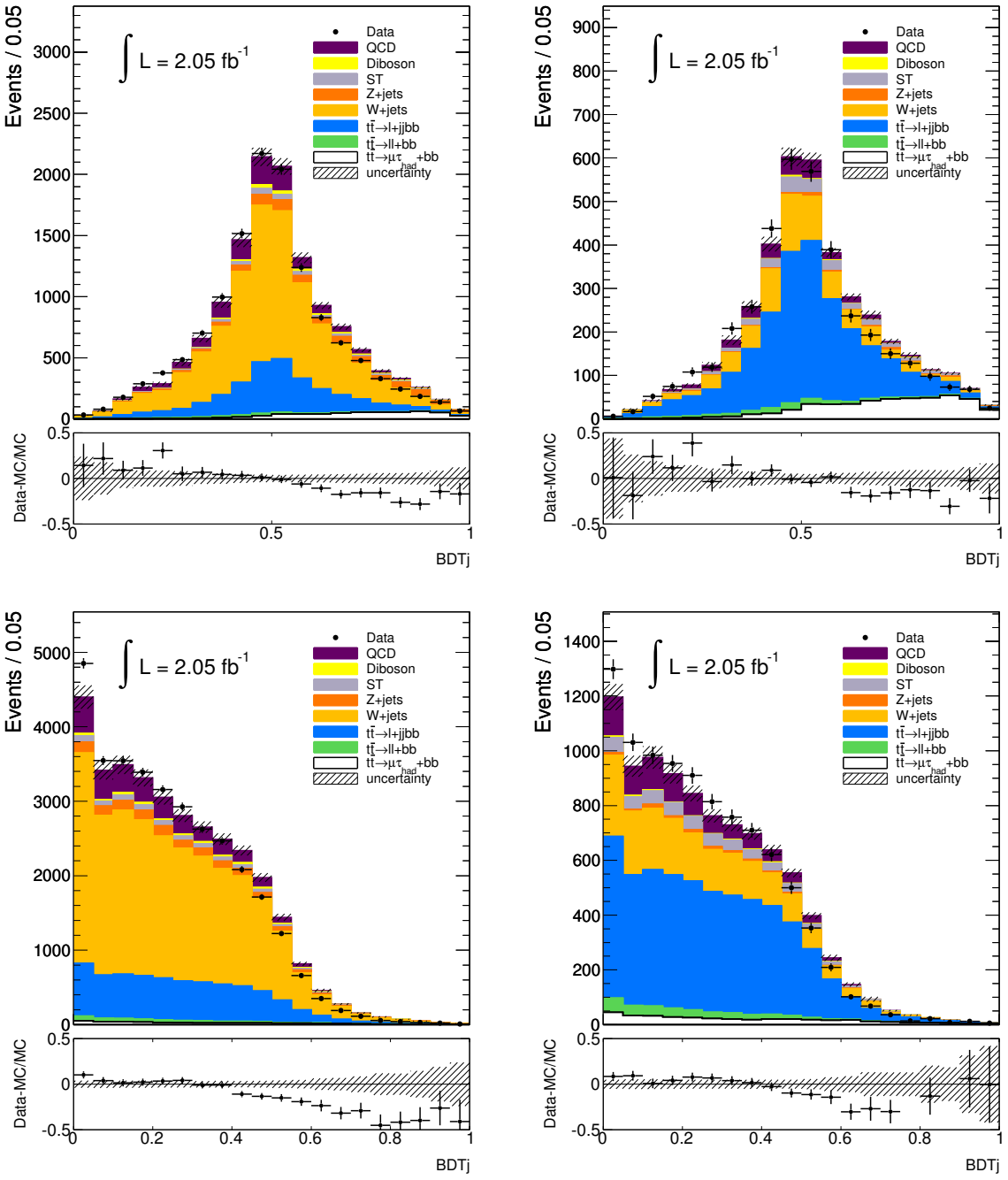
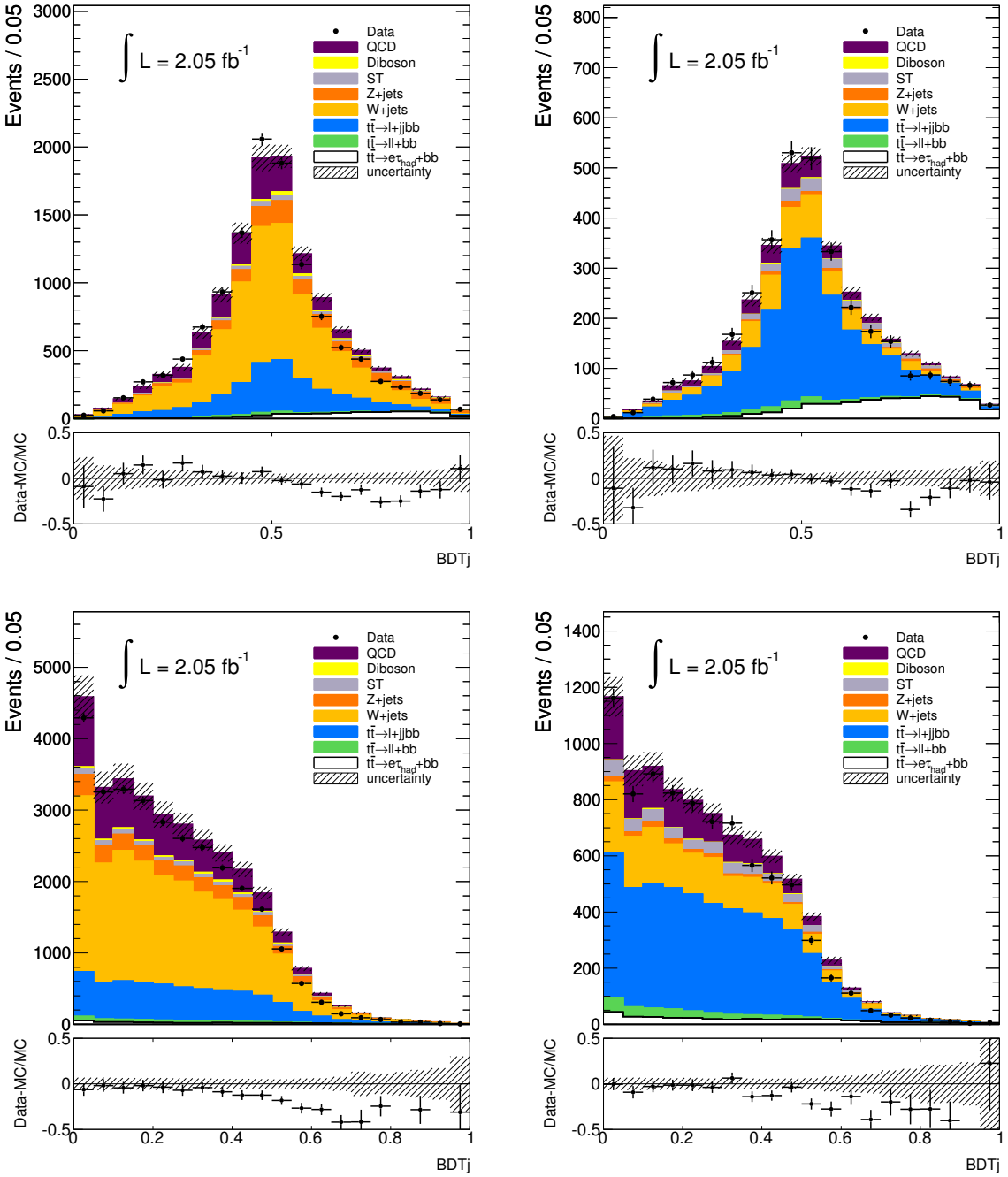


Figure 5.2: *BDT_j distributions of all τ_1 (top) and τ_3 (bottom) candidates in Data and MC for the e + τ channel. On the left plots the BDT_j distributions are shown after the events selections specified in Section 5.2 and before applying the b-tagging requirement. At this stage $t\bar{t} \rightarrow \ell$ + jet and W+ jets are the dominant backgrounds. After applying the b-tagging requirement the W+ jets background has been highly suppressed and $t\bar{t} \rightarrow \ell$ + jet becomes the dominant background, as shown in the right hand plots.*



5.3 Description of the fitting technique

After applying the event selections described in section 5.2, which include the kinematic selection cuts and the requirement of at least one jet passing the b-tagging identification, the dominant background is top quark pair production in the lepton plus jets channel. The only powerful suppression technique against this remaining background is tau identification. Therefore, an analysis technique to measure the $t\bar{t}$ cross section based on fits to distributions of signal and background templates will use the tau BDT_j variable to build these templates, thus exploiting the shape differences to discriminate the signal from the background processes enumerated in section 5.1.

There are various types of background:

1. from processes other than $t\bar{t}$ that have τ leptons and an isolated ℓ from W decay ($W \to \ell\nu$) or a tau leptonic decay ($\tau \to \ell\nu$). These processes include $Z \to \tau^+\tau^-$, single top Wt production ($W(\to \tau/\ell, \nu)t(\to W(\to \ell/\tau, \nu)b)$) and diboson events ($WW$, ZZ, WZ).

2. from processes with two isolated leptons where one lepton is misidentified as a τ. This type of background includes $t\bar{t}$ dilepton, $Z \to e^+e^-$, $Z \to \mu^+\mu^-$ and diboson processes.

3. from processes with a lepton where a jet is misidentified as a τ. $t\bar{t}$ semileptonic decays ($t\bar{t} \to \ell$ + jet), W+ jets and the remaining single top processes, s-channel and t-channel, contribute to this type of background.

4. from multi-jet processes where both lepton and τ are misidentified or the lepton comes from a jet not removed by the isolation requirement.

Figure 5.3 shows the BDT_j distribution of all MC tau candidates passing the event selection described in section 5.2 per jet type. Events where the lepton and tau have opposite sign charge (OS) are shown on the positive y axis and those with same sign charge (SS) leptons are shown in the negative y axis. We can see that τ leptons that are truth-matched to a real τ are OS while fake τ candidates contribute both to OS and SS events. The signal and background templates for the fit will be derived using OS τ candidates.

The signal BDT_j template is derived from MC τ candidates that are truth-matched to a real τ in the proportion expected from MC events passing the event selection. This selection includes the $t\bar{t} \to \ell + \tau$ signal, where the lepton comes from a W decay or a tau leptonic decay:

$$t\bar{t} \to W(\to \tau_{had}\nu_\tau)W(\to \ell\nu_\ell)b\bar{b} \tag{5.9}$$

$$t\bar{t} \to W(\to \tau_{had}\nu_\tau)W(\to \tau(\to \ell\bar{\nu}_\ell\nu_\tau)\nu_\tau)b\bar{b} \tag{5.10}$$

and non-$t\bar{t}$ processes like $Z \to \tau^+\tau^-$ and a small contribution from single top and diboson events. The main backgrounds of type 2 are $Z \to e^+e^-$ and $t\bar{t} \to \ell + e$ where an electron has been reconstructed as a τ candidate (mainly in the $e+\tau$ channel). Most electrons are removed by the BDT_e cut. For the few that remain, their BDT_j shapes are indistinguishable from τ leptons. Therefore, τ candidates that are truth-matched to real electrons are added to the signal template. The contributions to the signal template are shown in Table 5.6.

Backgrounds of type 3 and 4 are dominated by light-quark and gluon jets faking the τ candidate. These backgrounds will be modelled using templates built with data samples enriched in

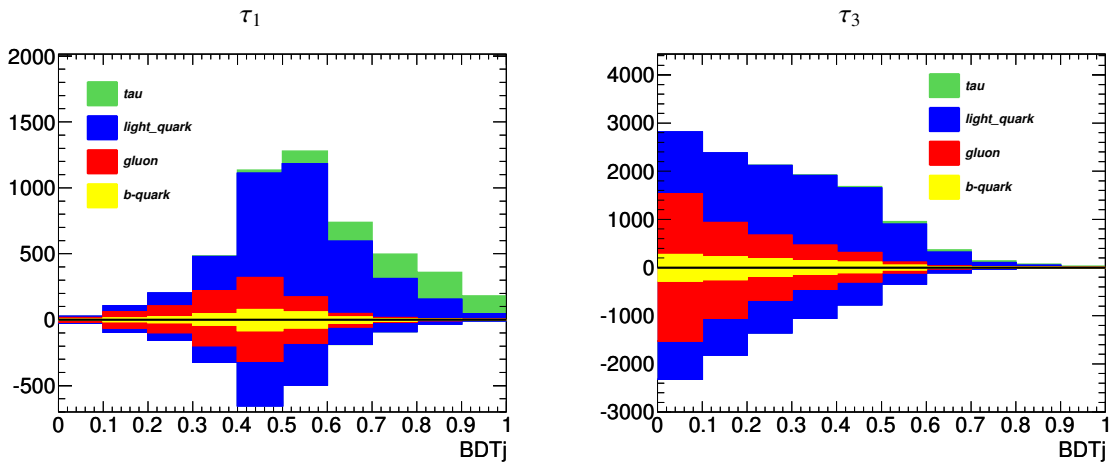


Figure 5.3: *BDT_j distribution by truth type of the reconstructed τ object. OS events are shown in the positive y axis and SS events are shown in the negative y axis. Real τ candidates contribute only to the OS events, while fake τ candidates contribute to both OS and SS events.*

	τ_1			τ_3		
	Total	e	τ	Total	e	τ
$t\bar{t}(\ell\tau_{had})$	740.4	0.06%	99.94%	236.2	0.05%	99.95%
$t\bar{t}(\ell\, jet)$	1.8	7.6%	92.4%	0.39	73.9%	26.1%
$t\bar{t}(\ell\ell)$	23.1	99.1%	0.8%	4.8	99.5%	0.5%
Z + jets	46.1	14.3%	85.7%	8.8	12.7%	87.3%
W + jets	0	-	-	0.6	0%	100%
Single Top	36.5	2.4%	97.6%	12.	1.5%	98.5%
Diboson	1.6	8.2%	91.8%	1.2	1.5%	98.5%
Total	849.6	3.7%	96.3%	264.1	2.5%	97.5%

Table 5.6: *Tau jet composition in MC signal template for each contributing channel.*

W+jets events. Within these samples, particular control regions are defined so that they contain different proportions of light-quark and gluon jets, as described in the Section 5.4. The fact that the background templates are derived from control regions in data allows us to avoid systematic effects related to jet composition in the MC models. Studies of the jet composition in the W+ 1 jet sample, shown in Appendix B, determine that the Monte Carlo simulation does not reproduce properly the BDT_j distribution in a data sample, thus reinforcing the decision to derive the background templates from data.

The background estimation and the signal measurement are established through fitting signal and background templates to the BDTj distribution in data. The new fitting technique aims to improve the background model of jets faking the hadronically decaying tau leptons by disen-

tangling the different shapes of the BDT_j distributions built with with light-quark and gluon jets.

Therefore, the contribution from $t\bar{t} \to \ell + \tau$ is derived from the $\geq 1b$-tag data sample, with the tau and the lepton oppositely charged, using a Chi-square (χ^2) fit to the BDT_j distribution with a signal template and two background templates: a gluon template and a light-quark template. The shapes of the templates are fixed, and the amount of signal and the two backgrounds, gluon and light-quark jets, separately, are the parameters of the fit. This is a two parameter fit after imposing the condition that the sum of signal and background must add up to the number of observed events. The fits are performed keeping τ_1 and τ_3 candidates separated, since the background level and BDT_j shapes are quite different.

5.4 Data-derived background templates

A crucial part of the new technique is obtaining the templates for background modelling. For that purpose we use W + jets selections from data, which contain gluon and light-quark jets in different proportions.

5.4.1 Gluon template extraction

We can extract the gluon template using a selection of W + jets from data:

- one isolated lepton (electron or muon) as described in 3.2
- only one reconstructed τ, after applying overlap removal with b-jets
- $E_{\mathrm{T}}^{\mathrm{miss}} \geq 30$ GeV to reduce the QCD multi-jet background
- $40 \leq M_{\mathrm{T}}(E_{\mathrm{T}}^{\mathrm{miss}}, \ell) \leq 90$ GeV to reject Z + jets events
- if there is only one jet in the event and it overlaps with the τ candidate the event will be classified as a W + 1 jet event. If there are two jets in the event and the τ candidate overlaps with the low pt one then the event will be considered a W + 2 jets event.

Both types of events are formed by different amounts of gluons and light-quark jets. The OS W + 1 jet sample is enriched in light quarks, while the SS W + 1 jet is a mixture of gluons and light-quark jets. In the W + 2 jets sample, the lower p_{T} jet has a high probability of being from initial or final state radiation and, therefore, a high probability of being a gluon jet. As the BDT_j distribution from gluon jets is charge symmetric, it is expected to have the same shape in SS and OS events, and contribute the same number of events to OS and SS for both the W + 1 jet and the W + 2 jets samples within the statistical precision:

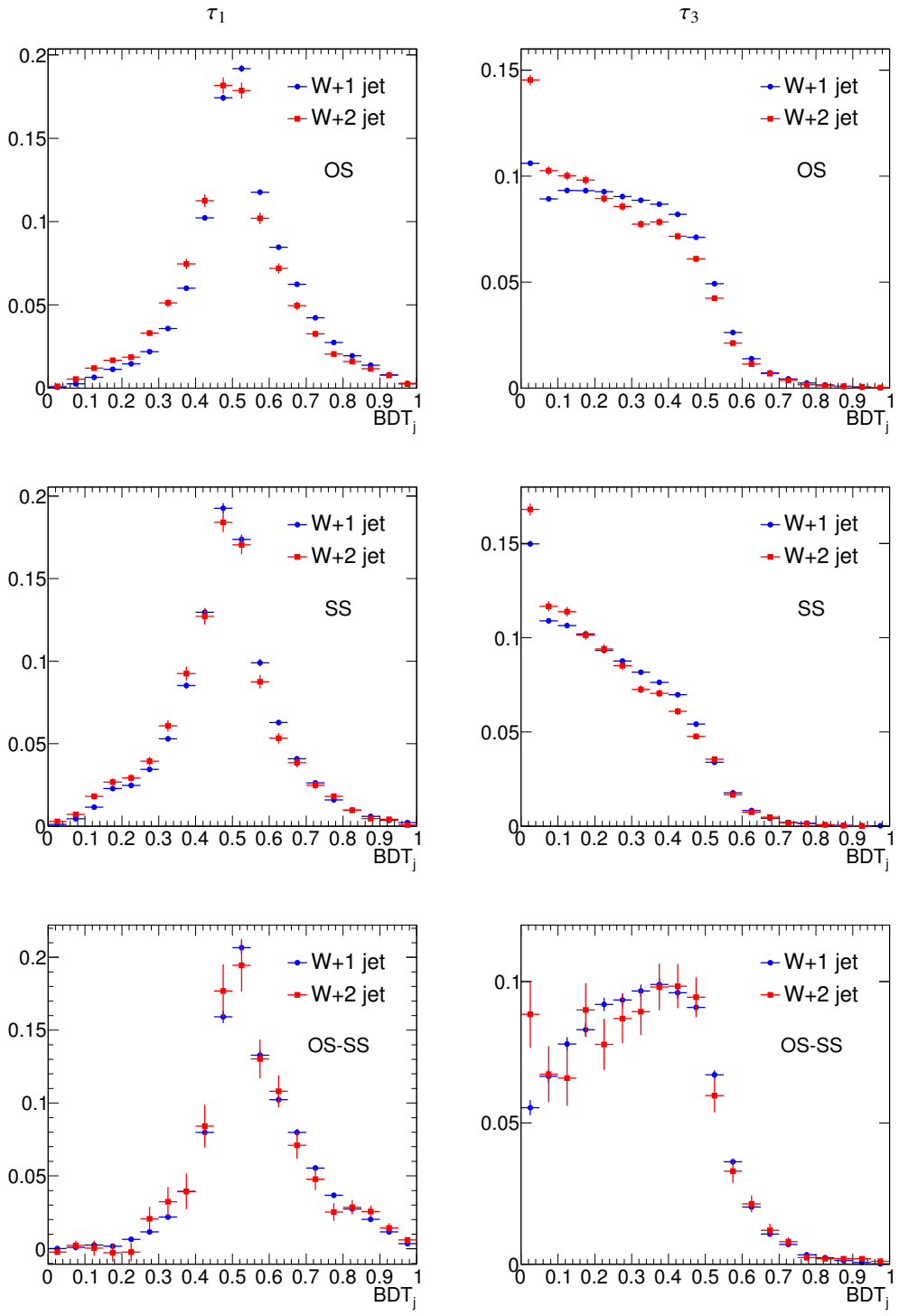
$$\mathrm{OS1} = a_1 \cdot \mathrm{OS_q} + b_1 \cdot \mathrm{Gluons}, \tag{5.11}$$

$$\mathrm{SS1} = c_1 \cdot \mathrm{SS_q} + b_1 \cdot \mathrm{Gluons}, \tag{5.12}$$

$$\mathrm{OS2} = a_2 \cdot \mathrm{OS_q} + b_2 \cdot \mathrm{Gluons}, \tag{5.13}$$

$$\mathrm{SS2} = c_2 \cdot \mathrm{SS_q} + b_2 \cdot \mathrm{Gluons}, \tag{5.14}$$

Figure 5.4: *OS (top), SS (middle) and OS-SS (bottom) BDT$_j$ distributions for τ candidates from W + 1 jet and W + 2 jets data samples. The left column is for τ_1 and the right column for τ_3.*



where OS1(SS1) and OS2(SS2) are the BDT_j distributions built from W + 1 jet and W + 2 jets samples, OS_q (SS_q) are the BDT_j distribution of light-quark jets and Gluons is the BDT_j distribution of gluon jets. In figure 5.4 we can see that there are significant differences between the shapes of the W + 1 jet OS(SS) and the W + 2 jets OS(SS) BDT_j distributions. However, if we substract the SS from the OS in both cases, the OS-SS distribution shapes are in very good agreement, which leads to the conclusions that $a_1/c_1 = a_2/c_2$ for any E_T^τ, as the transverse energy of τ candidates found in the W + 2 jets data samples is significantly lower than from W + 1 jet.

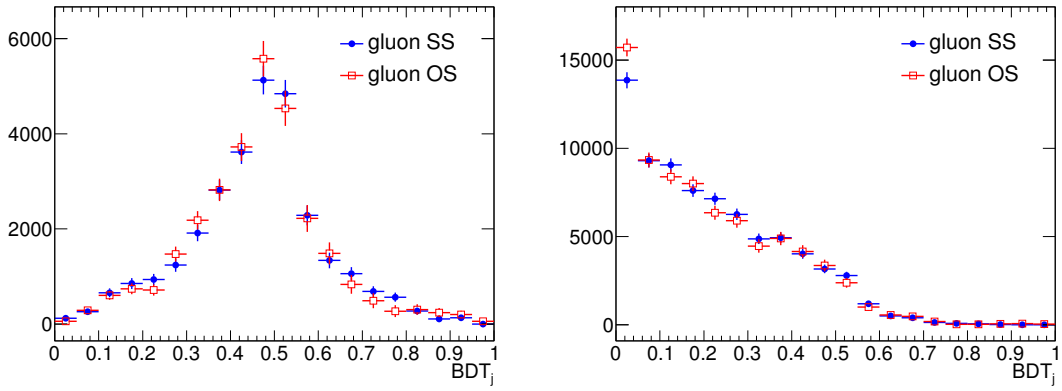
One can extract from the above equations the distribution Gluons from the OS and from the SS distributions separately:

$$K \cdot \text{Gluons} = (N \cdot \text{OS2} - \text{OS1}), \tag{5.15}$$

$$K \cdot \text{Gluons} = (N \cdot \text{SS2} - \text{SS1}), \tag{5.16}$$

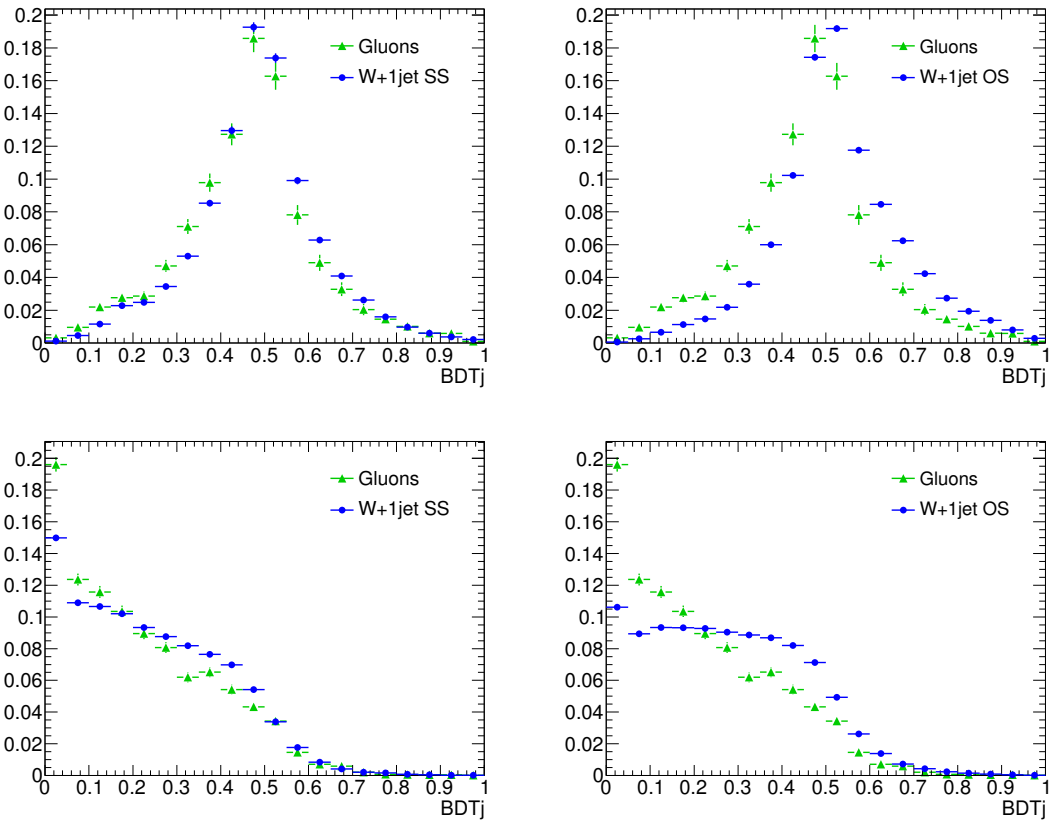
where N is the ratio of the integral of the subtracted distribution OS1-SS1 to the one from the distribution OS2-SS2 and $K = N \cdot b_2 - b_1$ is an unknown constant but it must be the same whether SS or OS is used to extract Gluons. Figure 5.5 shows both the extracted $K \cdot$ Gluons distribution through equations 5.15 and 5.16. It can be observed that they are in fact identical within statistics. Thus, they can be added together to reduce the statistical uncertainties.

Figure 5.5: *Extracted BDT_j gluon distributions for τ candidates from W + 1 jet and W + 2 jets samples. SS gluon distributions are extracted from SS W + 1 jet and SS W + 2 jets BDT_j distributions, while OS gluon distributions are extracted from OS W + 1 jet and OS W + 2 jets BDT_j distributions. The left plot is for τ_1 and the right one for τ_3.*



Figures 5.6 and 5.7 compare the shapes of the extracted gluon jet distributions to the W + 1 jet and W + 2 jets OS(SS) distributions. They show clearly that the W + 1 jet distributions, with a high percentage of fake τ candidates from light-quark jets, have higher BDT_j values than gluon jets. On the other hand, the W + 2 jets distributions are much closer to the extracted gluon distributions, as expected from the large percentage of gluons in the lower p_T jets from W + 2 jets.

Figure 5.6: *Extracted BDT_j gluon normalized distributions for τ candidates from W + 1 jet and W + 2 jets samples adding the gluon distributions derived from OS and from SS compared to OS1 and SS1 distributions. The top row are for τ_1 and bottom row for τ_3.*



5.4.2 Quark templates from a W+ jets selection

Unlike the gluon jet distributions, the OS_q and SS_q distributions can not be easily extracted from data, because we can not obtain the exact amount of the gluon jet contribution in equations 5.11 to 5.14.

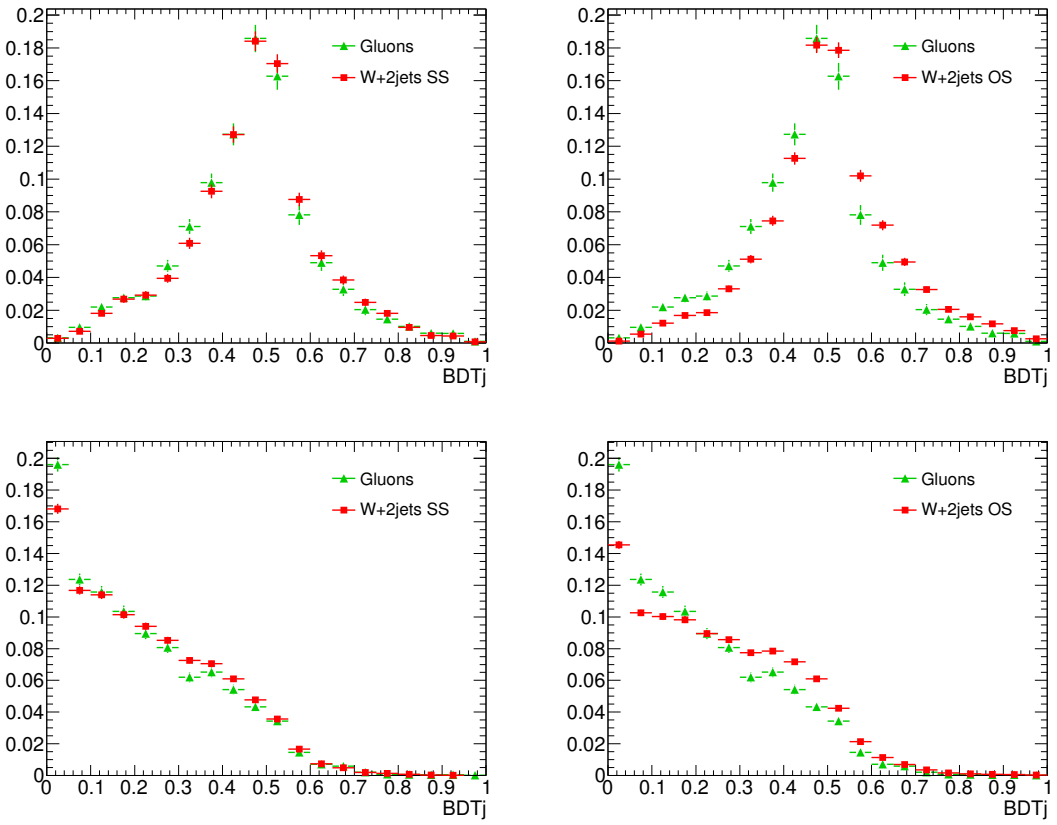
However, the BDT_j distribution modelling the background originated from jets faking tau leptons affecting the $t\bar{t} \rightarrow \ell + \tau$ signal can in principle be described by a combination of OS1(SS1) and Gluons:

$$\mathrm{OS(SS)} = \mathrm{k_1} \cdot \mathrm{OS1(SS1)} + \mathrm{k_2} \cdot \mathrm{Gluons}, \tag{5.17}$$

where k_2 can be negative if OS(SS) has a smaller percentage of gluon jets than OS1(SS1). The constants k_1 and k_2 can be obtained by fitting OS(SS) with OS1(SS1) and Gluons templates.

Given that the $t\bar{t} \geq 1$ b-tag data sample does contain real τ leptons, a third template built with

Figure 5.7: *Extracted BDT_j gluon normalized distributions for τ candidates from W + 1 jet and W + 2 jets samples adding the gluon distributions derived from OS and from SS compared to OS2 and SS2 distributions. The top row are for τ_1 and bottom row for τ_3.*



τ leptons needs to be included in the fit. In the case of $t\bar{t} \to \ell + \tau$, where the lepton, electron or muon, and the τ candidate have opposite charges, only the OS data will contain τ leptons while the SS events will be pure background:

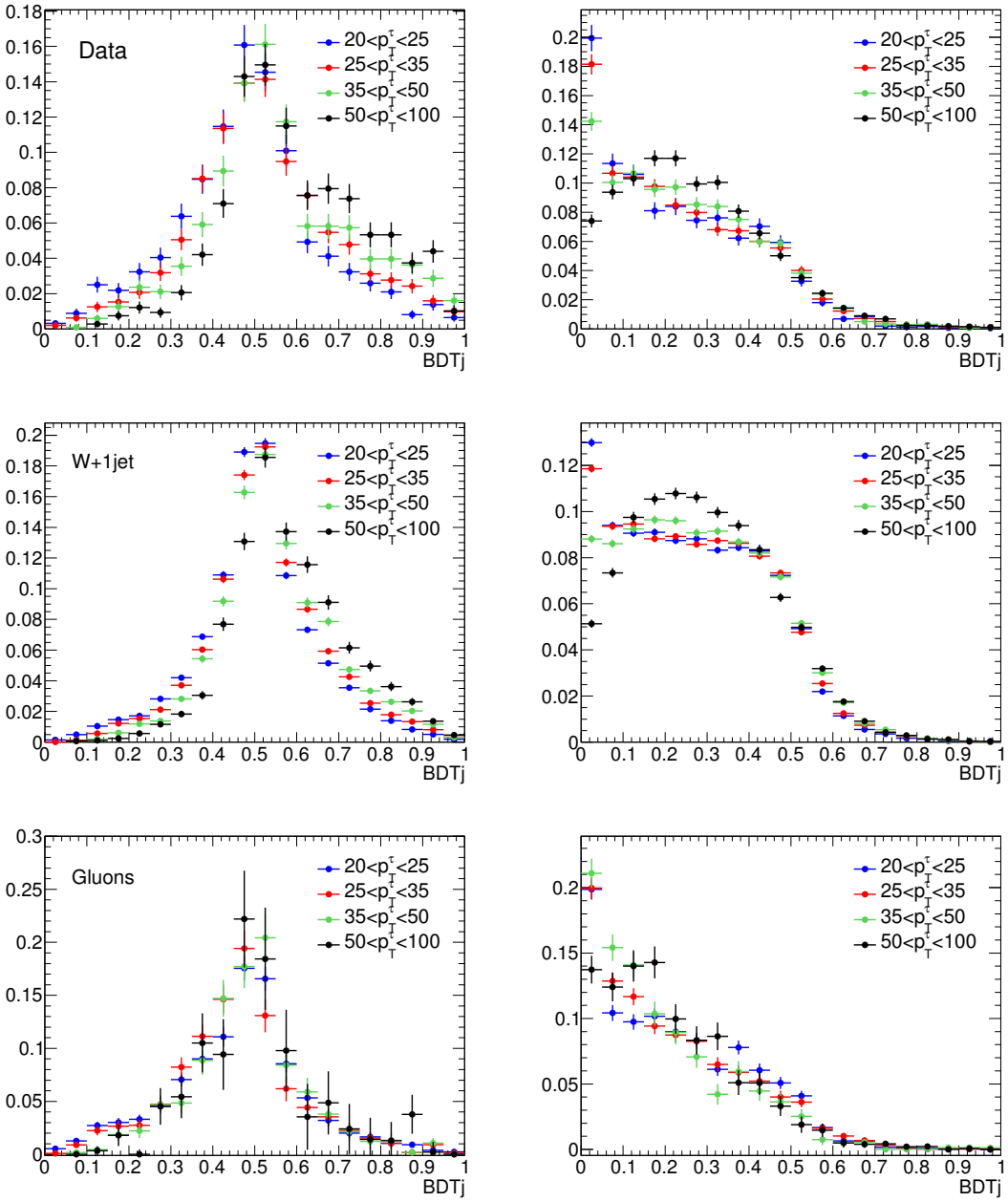
$$\mathrm{OS} = \mathrm{k_{OS1}} \cdot \mathrm{OS1} + \mathrm{k_g} \cdot \mathrm{Gluons} + \mathrm{k_s} \cdot \mathrm{Signal}, \tag{5.18}$$

$$\mathrm{SS} = \mathrm{k_{SS1}} \cdot \mathrm{SS1} + \mathrm{k_g} \cdot \mathrm{Gluons}. \tag{5.19}$$

Therefore, we can fit the BDT_j distribution for the $t\bar{t} \geq 1$ b-tag data sample with W + 1 jet and Gluon templates derived from data and a τ lepton signal template derived from MC to obtain the number of events with real τ leptons in the 2 fb^{-1} $t\bar{t}$ sample.

***BDT_j* distributions dependence on p_T of τ candidate.** The shape of the BDT_j distributions depends on the p_T of the τ candidates, as can be seen in figure 5.8, where the BDT_j distribution

Figure 5.8: *BDT_j distributions for τ candidates in data (top), W + 1 jet (middle) and gluon (bottom) samples after event selection. The BDT_j distributions are shown in 4 different p_T bins between 20 and 100 GeV to illustrate the BDT_j shape variation with the p_T of the τ candidate. The left plot is for τ_1 and the right one for τ_3.*



of $\tau_{1,3}$ candidates in data (top row), $W + 1$ jet (middle row) and gluon (bottom row) data samples are shown in 4 different p_{T} bins between 20 and 100 GeV. These bins have been selected so their population and their statistical errors are similar.

Therefore, the strategy we followed is to construct templates for the data, signal and background BDT_j distributions for different p_{T} ranges and do individual fits in each one. Due to the limited statistics in the 2 fb^{-1} sample, only two p_{T} bins are used in this analysis: $20 < p_{\mathrm{T}} < 35$ and $35 < p_{\mathrm{T}} < 100$ GeV.

Fits on $e + \tau$ and $\mu + \tau$ channels. The comparison of the BDT_j templates, displayed in Figure 5.9, shows that the shapes of the e and μ channels are in good agreement with each other, as was expected since the shape of the BDT_j of the τ candidates does not depend on the channel.

Taking into consideration the limited statistics in the 2 fb^{-1} sample, all templates are built by adding the BDT_j distributions for $e + \tau$ and $\mu + \tau$ channels in all data control regions. The fits are performed to the resulting templates and thus simultaneously for the two channels.

5.5 Signal extraction by fits to BDT_j distributions

The fitting of the data BDT_j distribution is done using Chi-squared minimization. The method of least squares is built on the hypothesis that the optimum description of a set of data is one which minimizes the weighted sum of squares of deviations, Δy, between the data in the i-bin of the BDT_j distribution, y_i, and the fitting function f_i:

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - f_i)^2}{\sigma_{y_i}^2 + \sigma_{f_i}^2}, \tag{5.20}$$

where N is the number of data points (number of bins in the BDT_j templates), f is the value of the signal and background combination at each bin:

$$f = k_s \cdot \text{signal} + k_g \cdot \text{Gluons} + k_q \cdot \text{OS}_{\text{q}}, \tag{5.21}$$

σ_{y_i} is the data error in the ith bin and σ_{f_i} is the the error of the function f obtained by propagating the errors of the signal, gluon and OS_q templates in the ith bin through the equation 5.21.

The optimal values of the normalization parameters k_g, k_q and k_s are found iteratively, minimizing the χ^2. This is a two parameter fit since the total number of background plus signal events is constrained by the number of events in the data sample.

The χ^2 minimization fits presented in the following sections have been performed using a numerical function minimization program widely used in particle physics: MINUIT [139], which is included in the ROOT framework.

5.5.1 Testing the fitting method with MC

Before applying the fitting procedure in data, the method was tested with MC to establish whether the method would extract the known signal in the sample. This is done in the following way:

Figure 5.9: *BDT_j distributions for τ candidates in data (top), W + 1 jet (middle) and gluon (bottom) samples after event selection for the $e + \tau$ and $\mu + \tau$ channels. The left plot is for τ_1 and the right one for τ_3.*

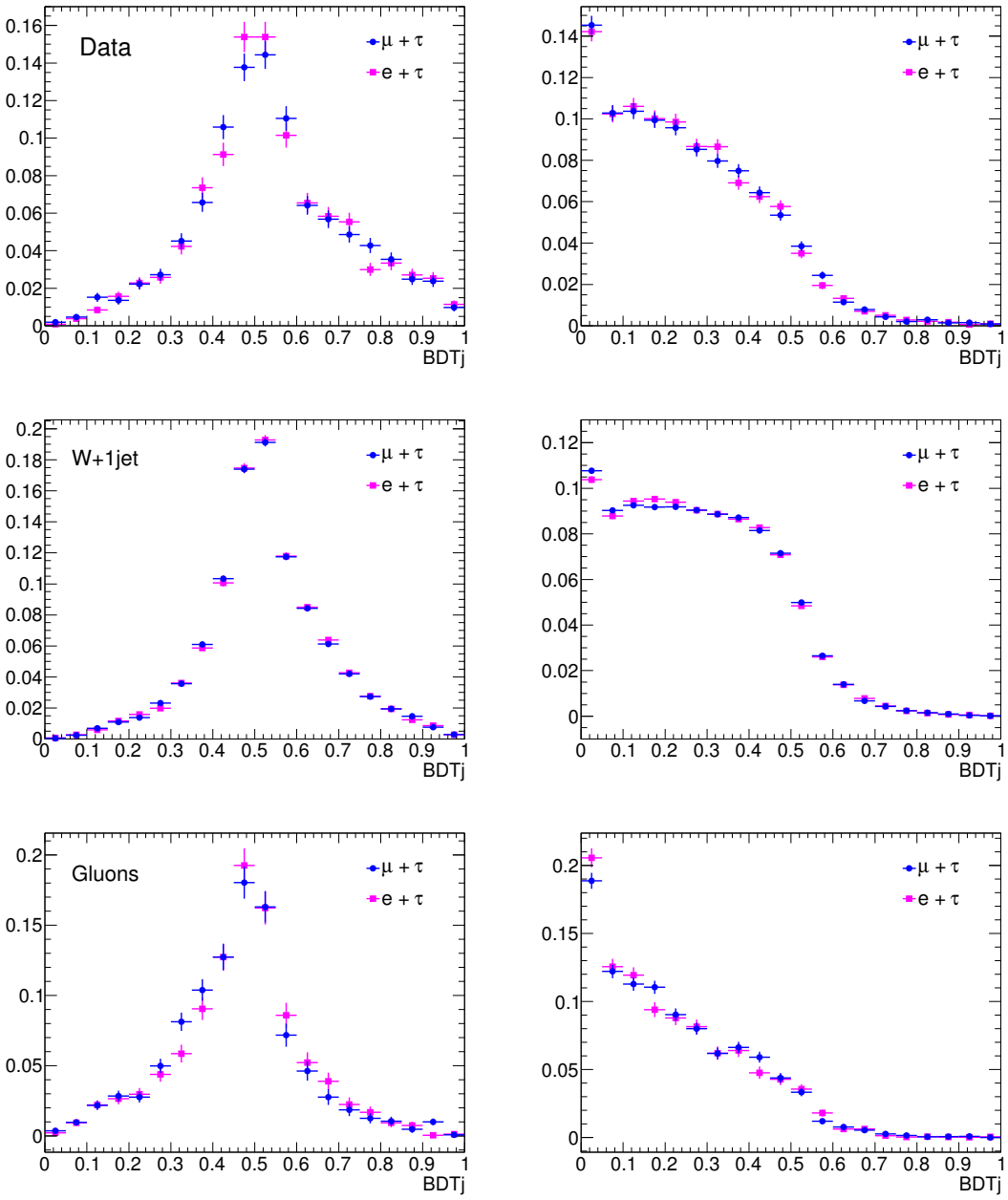
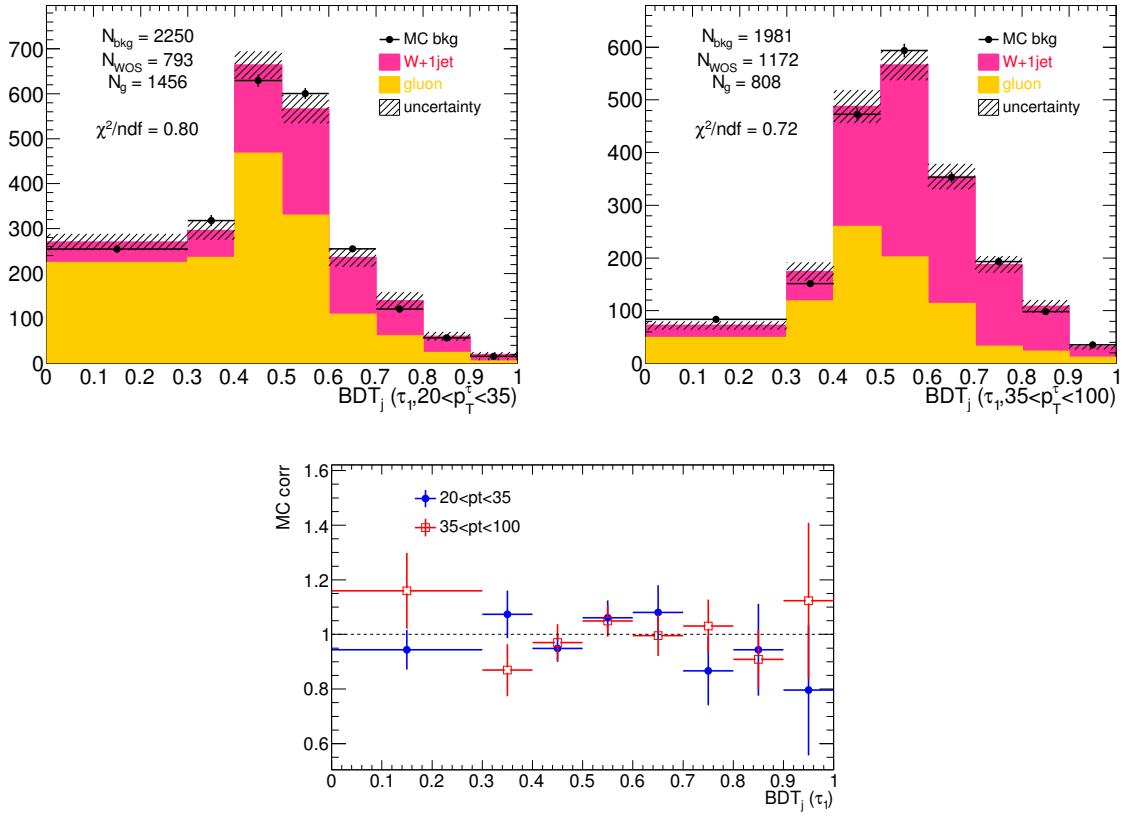
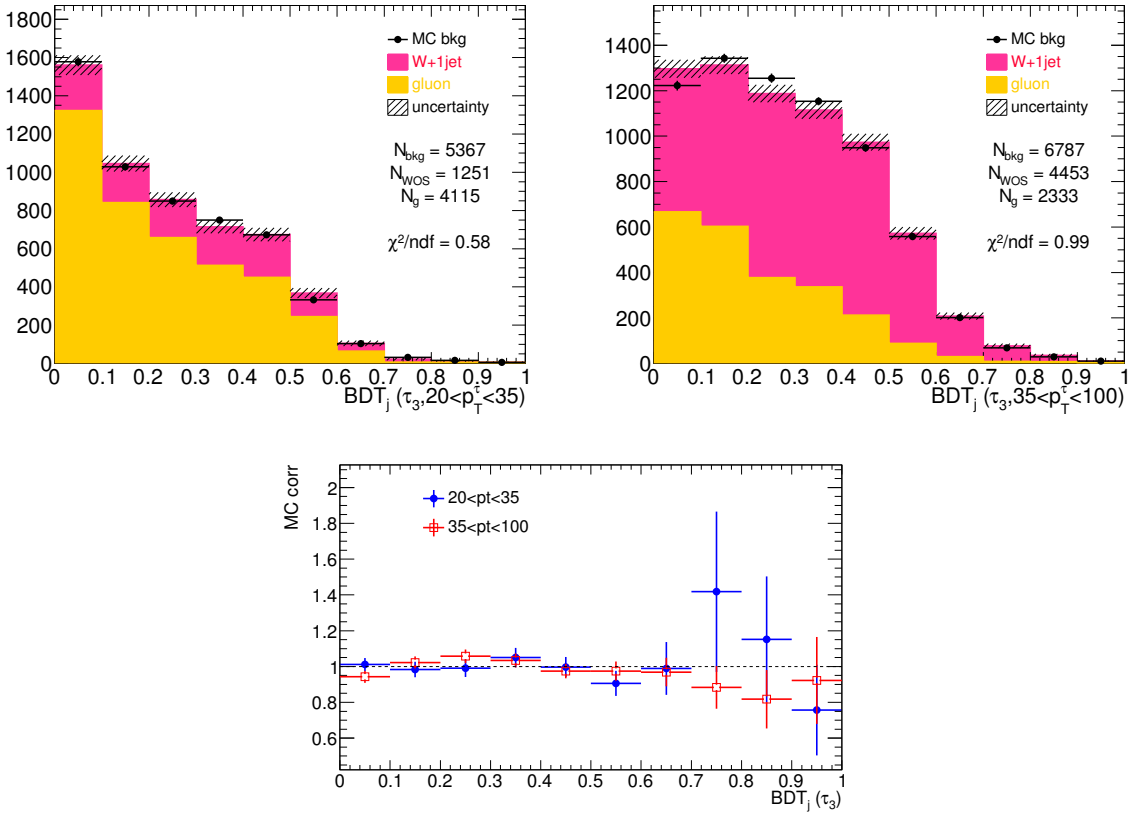


Figure 5.10: *Top: MC background BDT_j fit to W + 1 jet and gluon distributions for τ_1 in both pt bins. Left plot is for τ leptons with $20 \leq p_T \leq 35$ GeV. Right plot shows τ leptons with $35 < p_T \leq 100$ GeV. Bottom: BDT_j background correction factors derived from above fits.*



- First we test how accurate the background description is by fitting the MC background with the gluon and quark templates. The MC background BDT_j templates are built using all τ candidates from the MC samples ($t\bar{t}$, W + jets, Z + jets, single top and dibosons) that are not included in the signal template, that is, all τ candidates that are not truth matched to a real τ or an electron. The QCD multi-jet background is not included in the MC background templates. The results show that the MC background is quite well described by the mixing of light-quarks and gluons, as can be seen in the top plots in figures 5.10 (τ_1) and 5.11 (τ_3). In each plot the orange histogram corresponds to the gluon template scaled by the k_g fit parameter, the pink histogram correspond to the sum of the gluon and W + 1 jet templates, both scaled by the fit parameters k_g and k_{qos}, and the solid points correspond to the MC background template. The dashed lines show the fit uncertainty. The reduced χ^2 of the fits are close to 1, ensuring the validity of the background description using light-quark and gluon jets.

Figure 5.11: *Top: MC background BDT_j fit to W + 1 jet and gluon distributions for τ_3 in both pt bins. Left plot is for τ leptons with $20 \leq p_T \leq 35$ GeV. Right plot shows τ leptons with $35 < p_T \leq 100$ GeV. Bottom: BDT_j background correction factors derived from above fits.*



- Shape differences between the actual and the fitted background templates can be corrected with a reweighting function equal to the ratio of both BDT_j distributions, shown in the lower plot of figures 5.10 (τ_1) and 5.11 (τ_3). The high uncertainty in some p_T bins can be accounted for by the lack of statistics in the MC samples.

- The next step is to fit the complete MC sample with the background and signal templates. If the background templates reproduce the actual background shape accurately enough we should recover the signal template without any significant bias. Figures 5.12 and 5.13 show the fitting results for τ_1 and τ_3 candidates in both pt bins for the original background templates (left) as well as the derived templates after applying the MC corrections (right). The number of fitted signal events obtained without applying the MC corrections are 786 for the τ_1 and 218 for the τ_3. These represent a deviation of 9% and 14%, respectively, from the expected number of signal events which are 849 for the τ_1 and 264 for the τ_3. In the fits performed with the MC corrected background templates, no signal bias is

observed. This improvement justifies the decision to apply the MC correction functions to the data derived background templates for the signal extraction in the 2 fb^{-1} data sample. The MC fitting results in both p_T bins are shown in table 5.7. The uncertainties shown are from the fit only and do not include systematic uncertainties.

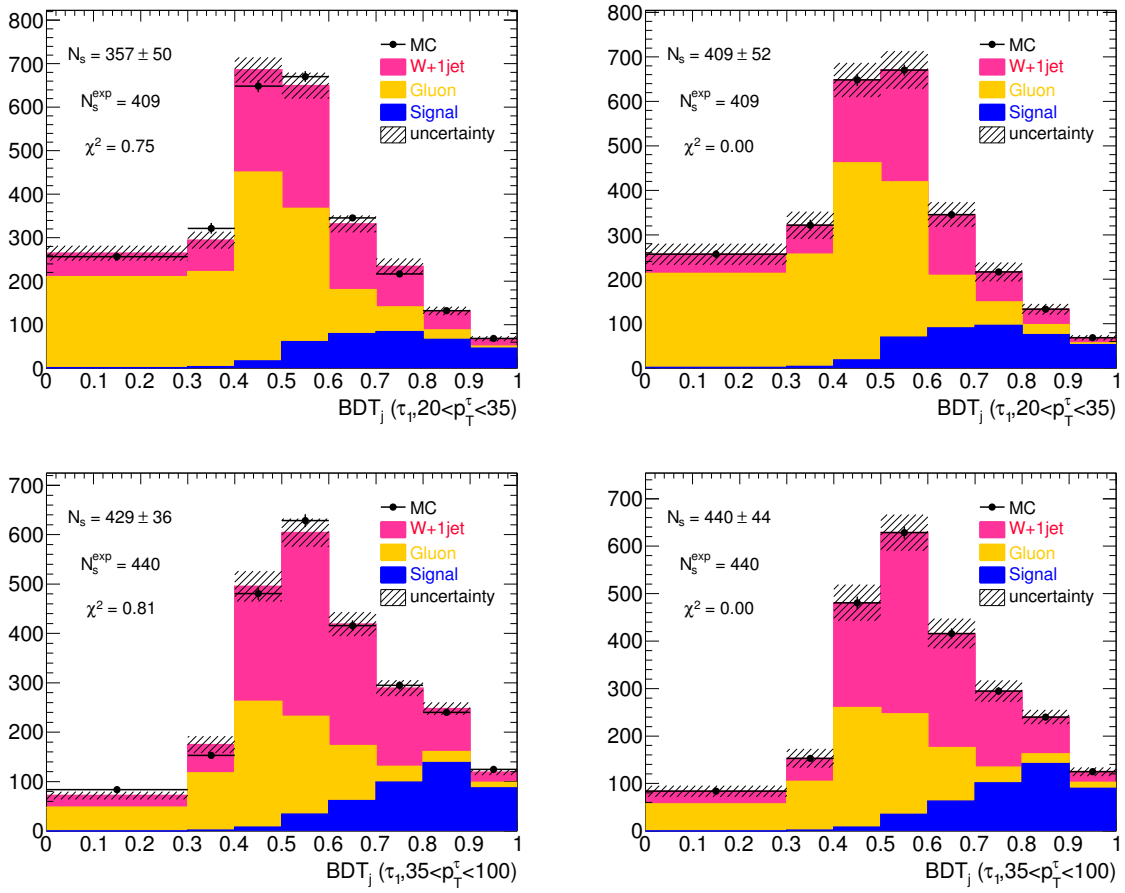


Figure 5.12: *MC Fits using original (left) and corrected (right) background templates for τ_1 candidates. Top row shows τ leptons with $20 \leq p_T \leq 35$ GeV. Lower row shows τ leptons with $35 < p_T \leq 100$ GeV.*

5.5.2 Applying the fitting method to data

After performing the MC test described in the previous section, the fitting method was applied to the 2.05 fb^{-1} ATLAS data sample in order to extract the $t\bar{t} \rightarrow \ell + \tau$ signal. The results are summarized at Table 5.8. In figure 5.14 we can see the data fits in both p_T bins for τ_1 (top) and τ_3 (bottom). The signal template scaled by the fitting parameter k_s is shown in blue, gluons in

MC Expectation	$20 < p_T < 35$ GeV	$35 < p_T < 100$ GeV	Total
τ_1	409 ± 6	440 ± 6	849 ± 8
τ_3	112 ± 2	152 ± 4	264 ± 4
Fit wo MC corrections			
τ_1	357 ± 50	429 ± 37	786 ± 62
τ_3	107 ± 30	111 ± 28	218 ± 41
Fit with MC corrections			
τ_1	409 ± 52	440 ± 44	849 ± 68
τ_3	112 ± 32	152 ± 32	264 ± 45

Table 5.7: *Number of τ leptons in the signal template in different p_T regions (top). The results from the fits performed with MC derived templates are shown with (bottom) and without (middle) applying the corrections to the background templates.*

orange and W + 1 jet in pink. The solid points correspond to the data and the dash lines are the fit statistical uncertainties.

5.5.2.1 Combined Fits in p_T bins.

A combined fit in the two p_T regions ($20 < p_T < 35$ GeV and $35 < p_T < 100$ GeV) has allowed to reduce substantially the statistical error on the cross section measurement. Instead of performing two independent fits, with a total of 4 free parameters, we do a single three parameter fit after imposing the conditions that the sum of signal and background must add up to the number of observed events in each p_T region. Only one signal parameter is used, weighted with the relative signal acceptance in each region as predicted by the MC.

It is a χ^2 minimization as described by equation 5.20 with a fitting function for each p_T bin described by:

$$f_j = k_s \cdot \frac{N^j_{\text{signal}}}{N_{\text{signal}}} \cdot \text{signal} + k_{gj} \cdot \text{Gluons} + k_{qj} \cdot \text{OS}_\text{q}, \tag{5.22}$$

where N_{signal} is the total number of events in the signal template for the whole p_T range and N^j_{signal} the number of events in each p_T bin signal template. k_s, k_{g1} and k_{g2} are the free parameters of the fit and k_{q1} and k_{q2} are determined by the total number of events in the data sample.

The results of these combined fits are shown in Figure 5.15 and summarized in Table 5.8. The signal extracted from the combined fit is included in the uncertainty of the individual p_T bin fits and the uncertainty has been reduced a 23% in the 1-prong case and 64% in the multi-prong case.

5.6 Systematic Uncertainties affecting the cross section measurement

The cross section measurement is affected by systematic uncertainties, originating from various sources. There are detector-related uncertainties, such as those on particle identification

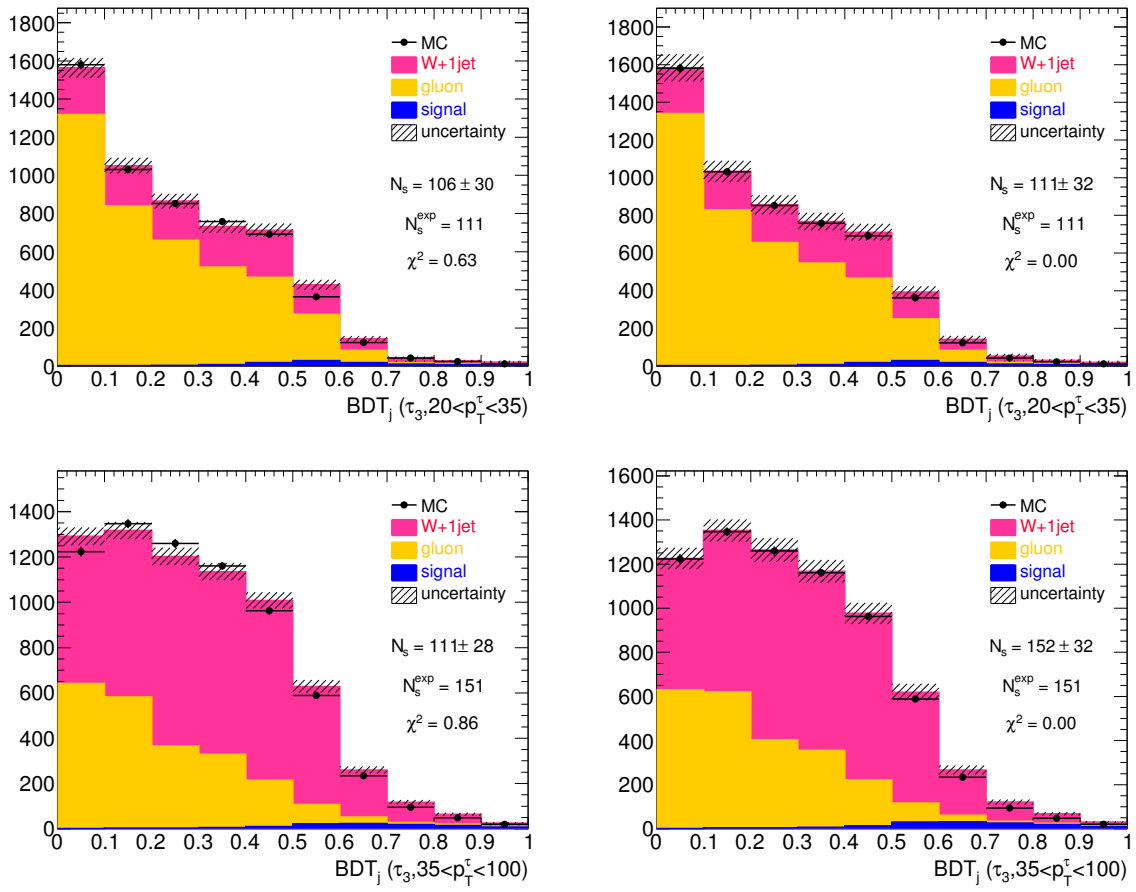


Figure 5.13: *MC Fits using original (left) and corrected (right) background templates for τ_3 candidates. Top row shows τ leptons with $20 \leq p_T \leq 35$ GeV. Lower row shows τ leptons with $35 < p_T \leq 100$ GeV.*

	$20 < p_T < 35$ GeV	$35 < p_T < 100$ GeV	Merged p_T bins	Combined p_T Fits	MC
τ_1	390 ± 49 (0.24)	416 ± 45 (0.56)	807 ± 67	802 ± 50 (0.39)	848 ± 8
τ_3	152 ± 84 (0.41)	127 ± 30 (0.59)	280 ± 89	232 ± 32 (0.68)	262 ± 4

Table 5.8: *Data results of template fits to $\ell + \tau$ BDT$_j$ distributions. Second and third columns show the fit results in the two p_T bins. The fourth column shows the total extracted signal. Numbers in the parentheses are the reduced χ^2. The fifth column are the results form the combined p_T fits. The MC expectation is the number of events expected from the OS signal template asuming the theoretical $t\bar{t}$ cross section (164 pb).*

efficiencies, on background rejections, and on the precise knowledge of energy scales and resolution functions. Other uncertainties come from the approximations made in Monte Carlo generators, modelling and from the theoretical calculation of cross sections.

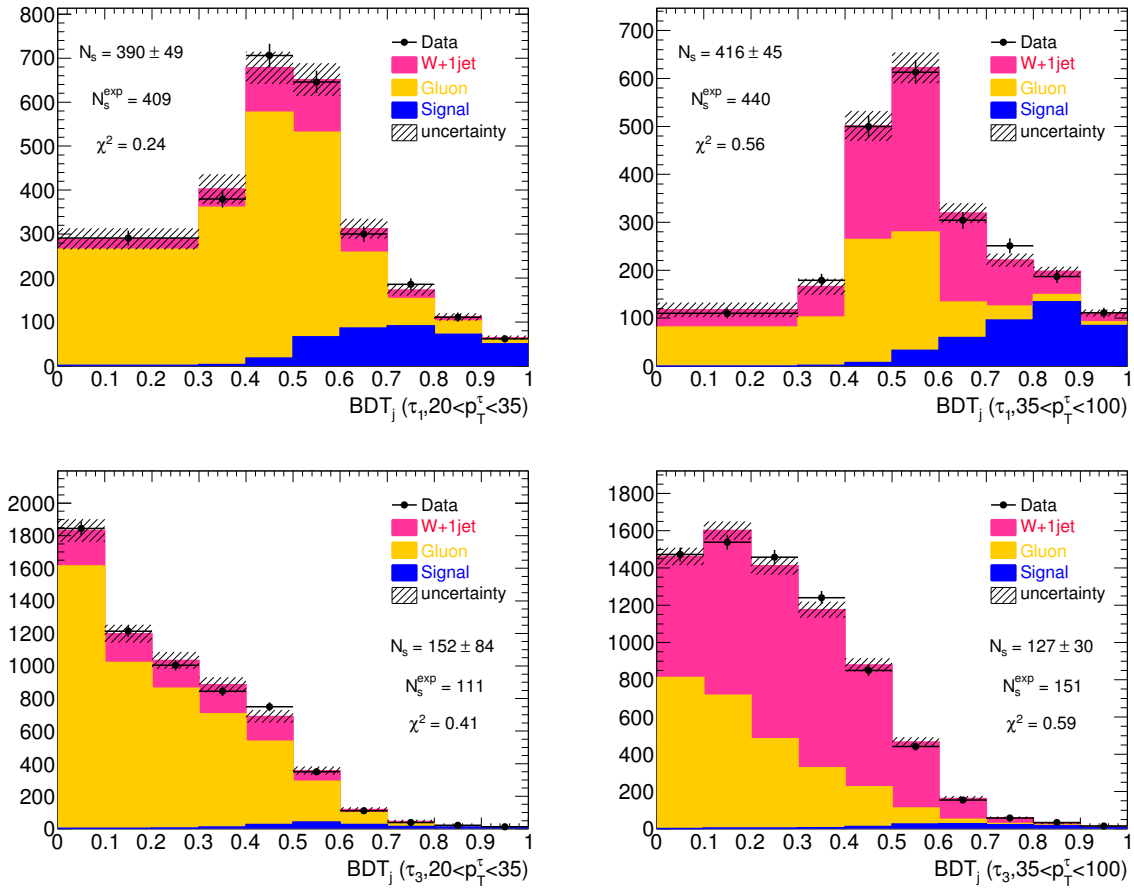


Figure 5.14: *Data Fits using MC corrected background templates for τ_1 (top) and τ_3 (bottom) candidates. Left column shows τ leptons with $20 \leq p_T \leq 35$ GeV. Right colum shows τ leptons with $35 < p_T \leq 100$ GeV.*

5.6.1 Uncertainties related to MC simulation

The uncertainty in the kinematic distributions of the $t\bar{t}$ signal results in systematic uncertainties in the signal acceptance. The main contributions are the choice of generator, the modelling of initial (ISR) and final (FSR) state radiation and the choice of the PDF set:

- **Monte Carlo Modelling of the Signal**. The uncertainty due to the use of different Monte Carlo generators is evaluated for the $t\bar{t}$ lepton plus tau signal by comparing the acceptance obtained with two different generators, MC@NLO and POWHEG, both interfaced to HERWIG. The parton shower systematic uncertainty on the $t\bar{t}$ acceptance is evaluated by comparing predictions of POWHEG interfaced to HERWIG to those of POWHEG interfaced to PYTHIA [140–142].

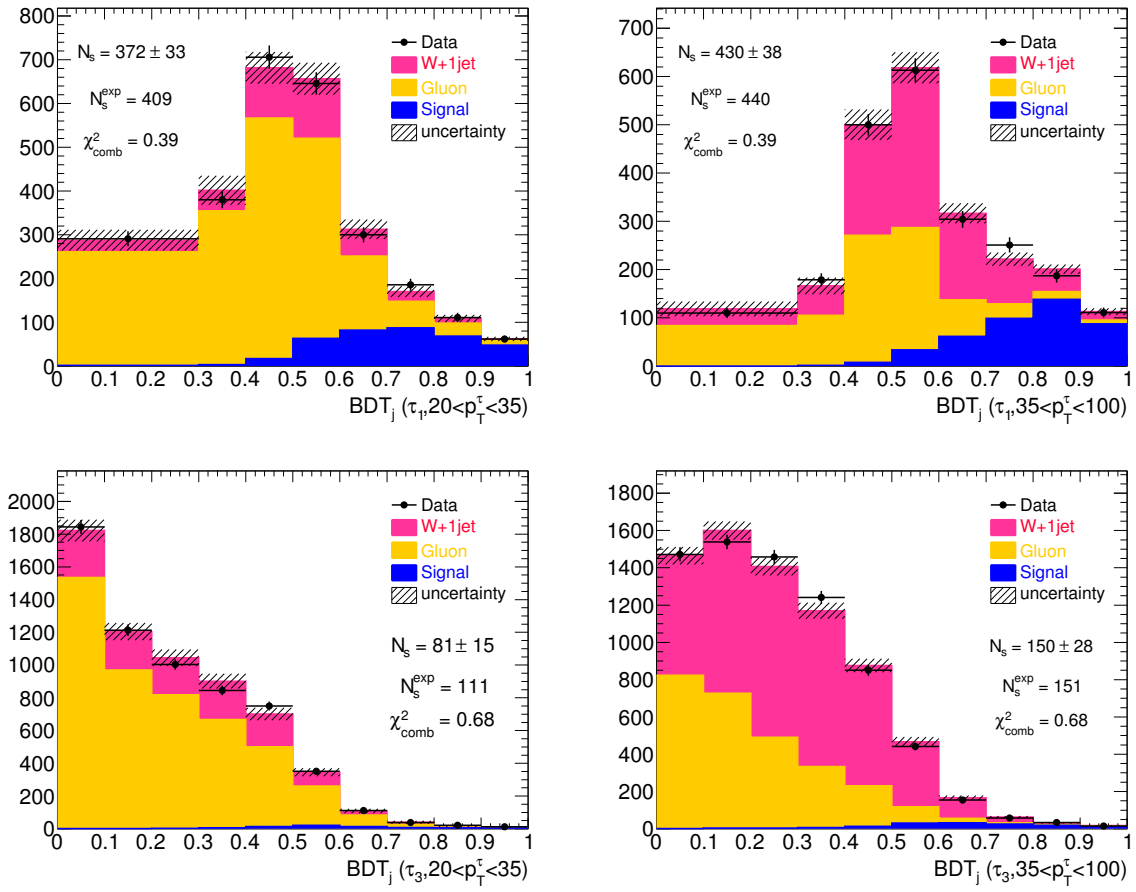


Figure 5.15: *Data Fits in the combined p_T bins using MC corrected background templates for τ_1 (top) and τ_3 (bottom) candidates. Left column shows τ leptons with $20 \leq p_T \leq 35$ GeV. Right colum shows τ leptons with $35 < p_T \leq 100$ GeV.*

- **PDF**: the effect of uncertainties in the parton density functions (PDF) is evaluated using a range of current PDF sets: the standard CTEQ and two alternative ones: NNPDF and MSTW [143–145]. Each one comes with a set of error PDFs. The RMS of the variations was taken as the PDF uncertainty.

- **Initial and Final State Radiation**. More (Less) initial and final state QCD radiation (ISR and FSR) increases (decreases) the number of jets and affects the transverse momentum of particles in the event. The effect of ISR/FSR variation on signal acceptance is studied using the ACERMC generator interfaced to PYTHIA shower model [146], and varying the parameters controlling ISR and FSR in a range consistent with experimental data [95]. The largest relative differences with respect to the reference sample are used as systematic uncertainties.

The dominant uncertainties in this category of systematic uncertainties are the modelling of ISR and FSR.

5.6.2 Uncertainties related to detector performance

Other sources that affect the signal acceptance are the reconstruction and identification efficiencies, as well as the energy/momentum scale and resolution for the objects described in Section 3.2:

- **Lepton Reconstruction and Identification Efficiency, Energy Scale and Resolution**. Lepton trigger, reconstruction and selection efficiencies are assesed by comparing $Z/\gamma \rightarrow \ell^+\ell^-$ events selected with the same object criteria as used for the $t\bar{t}$ analysis in data and MC. To account for any differences in acceptance between data and MC, scale factors are applied to MC samples. These scale factors are obtained comparing observed and predicted selection efficiencies with $Z/\gamma \rightarrow \ell^+\ell^-$ events. Systematic uncertainties on this scale factors are evaluated by varying the selection of events used in the efficiency measurements and by checking the stability of the measurements over the course of data taking.

  The modelling of the lepton momentum scale and resolution is studied using reconstructed invariant mass distributions of $Z/\gamma \rightarrow \ell^+\ell^-$ candidate events and used to adjust the simulation accordingly (see references [147] and [148]). The energy scale is corrected as a function of η of the lepton. Systematic uncertainties are dominated by uncertainties from the the detector material and the presampler energy scale, but also include the event selection, pileup, and hardware modelling.

  The acceptance uncertainty from the lepton modelling is dominated mostly by the muon trigger uncertainty and the electron selection efficiency uncertainty.

- **Jet related uncertainties**.
  The Jet Energy Scale (JES) calibrates the measured calorimeter-level jet energy to the particle-level, taking into account the effect from neutrons, dead materials, other detector effects, and algorithm specific biases. The JES and its uncertainty are derived by combining information from test-beam data, LHC collision data and simulation [149, 150]. JES uncertainty varies in the range 4-8% as a function of jet p_{T} and η. The uncertainty due to the limited knowledge of the JES is determined by varying the energy of the reconstructed jets according to the estimated uncertainties. $E_{\mathrm{T}}^{\mathrm{miss}}$ is also rescaled accordingly in the shifted samples for the JES uncertainty.

  The jet reconstruction efficiency is defined as the fraction of jets built from charged tracks reconstructed in the inner detector system matched to a calorimeter jet. The jet energy resolution and jet reconstruction/identification efficiency measured in data and in simulation are in good agreement. The statistical uncertainties on the comparisons, 10% and 1-2% for the jet energy resolution (JER) and jet reconstruction efficiency (JetEff), respectively, are taken as systematic uncertainties associated with these effects.

- $E_{\mathrm{T}}^{\mathrm{miss}}$ **uncertainties**. The most significant sources of uncertainty related to the $E_{\mathrm{T}}^{\mathrm{miss}}$ come from the scale and resolution of the objects (electrons, jets and muons), the description of the pileup events, and the impact of hardware failures [117]. Each of the objects in

the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation has an uncertainty related to the scale and resolution of the energy and p_{T} of the object that is propagated into the $E_{\mathrm{T}}^{\mathrm{miss}}$. For high p_{T} jets the jet efficiency uncertainty is also taken into account.

- **B-tagging**. The uncertainty on the efficiency of the b-tagging algorithm has been estimated to be 6% for b-quark jets, based on b-tagging calibration studies using inclusive lepton and multi-jet final states [120].

- **Tau identification systematics** are derived from a template fit to a $Z \to \tau^+\tau^-$ data sample selected with the same object selection as the data sample for this analysis, but different kinematic requirements to enhance the Z + jets selection and remove W + jets events. A detailed derivation of the tau identification systematics can be found in Appendix C.

The dominant systematics in this category are τ lepton identification, b-tagging and JES, whereas JER and JetEff have a minor impact on the cross section measurement.

5.6.3 Uncertainties related to luminosity

- As we saw in Section 2.2.2, the measured luminosity has an uncertainty of 3.7% [83–85]. The effect of the luminosity variation directly affects to the signal acceptance, as well as the subtraction of the real τ and electron contributions. This translates into a 3.5% uncertainty on the cross section.

- **Pile Up**. The LHC instantaneous luminosity varied by several orders of magnitude during the 2011 data-taking period analysed in this thesis (see Section 3.1 for more details), reaching a peak of about $\mathcal{L}_{\mathrm{ins}} = 3.65 \cdot 10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$. At this luminosity an average of about 7 extra pp interactions are superimposed on each collision event. This pileup background produces additional activity in the detector, thus affecting the different objects (e, μ, jets...) reconstruction variables. The effects on the $t\bar{t}$ signal are assessed using simulation samples where additional pileup events have been overlayed before digitization and reconstruction. For the 2.05 fb^{-1} sample the impact of pileup in the cross section uncertainty is about 0.5%, but is not itemized in Table 5.10 since it is included in the tau identification systematic.

5.6.4 Systematic Uncertainties effect on the cross section measurement

The above described systematic uncertainties affect the cross section measurement through:

- MC expectation on the signal acceptance

- Shape distortion of the signal template

Therefore, the effect of these systematic variations on the final cross section result is evaluated as follows:

1. The signal sample is processed with each systematic source shifted by $\pm\sigma$.

2. Each of the shifted sources is used to derive a new signal template.

Table 5.9: *Systematic uncertainties for signal acceptance ($\Delta\mathcal{A}/\mathcal{A}$) in the $\mu + \tau$ and $e + \tau$ channels in %.*

	$\Delta\mathcal{A}/\mathcal{A}$(%)	
	τ_1	τ_3
Muon p_{T} smearing (ID)	±0.0	0.0 / +0.1
Muon p_{T} smearing (MS)	±0.0	±0.0
Muon Trigger SF	±1.5	±1.5
Muon ID SF	±0.0	±0.0
electron p_{T} smearing	±0.2	−0.2 / +0.3
electron energy scale	±0.3	−0.2 / +0.3
electron Trigger SF	±0.2	±0.2
electron ID SF	±1.2	±1.2
Jet energy scale	−2.7 / +2.3	−2.9 / +2.6
Jet energy resolution	±0.2	±0.05
Jet ID efficiency	0.0	0.0
b−tag SF	−5.7 / +5.0	−6.0 / +5.4
ISR/FSR	±3.5	±5.9
Parton shower	±4.0	±0.2
MC generator	±2.7	±1.2

3. The change in acceptance is evaluated in the signal region for every shifted signal template, comparing the difference in acceptance to the nominal signal template. The resulting systematic uncertainties for the signal acceptance are shown in table 5.9 for both the $\mu + \tau$ and $e + \tau$ channels.

4. For those systematic samples obtained from MC@NLO $t\bar{t}$ sample (Jet and lepton related systematics, including b-tagging) we evaluate the effect of the shape change in the signal template performing the fit again for each shifted signal template. We then calculate the effect on the final cross section uncertainty via error propagation on equation 5.27:

$$\varepsilon = \frac{\sigma_{t\bar{t}(SM)}}{S_{MC}} \sqrt{(\Delta S_{meas})^2 + \frac{S^2_{meas}}{S^2_{MC}}(\Delta S_{MC})^2}, \tag{5.23}$$

where ΔS_{meas} is the variation in the number of measured signal events when we use the shifted signal template for the fit instead of the nominal one of the analysis and ΔS_{MC} is the variation in the number of expected signal events. The actual impact of jet and lepton uncertainties in the shape of the BDT_j of the τ candidates is minimal. The difference in the signal measured with each sample, ΔS_{meas}, is negligible when compared with the acceptance uncertainty (and, therefore, ΔS_{MC}) which is the dominant component of the cross section systematic uncertainty.

5. For the rest of the systematic samples (PDF, MC generator, ISR/FSR) we rely on the acceptance uncertainty and do not redo the fit:

$$\varepsilon = \frac{\sigma_{t\bar{t}(SM)} \cdot S_{meas}}{S^2_{MC}} \cdot \Delta S_{MC}. \tag{5.24}$$

The reason to do so is that changing the MC generator of the $t\bar{t}$ process affects the MC corrections we apply on the data derived backgrounds, therefore changing the background templates for the fit and thus modifying the background model.

The effect of each systematic on the total cross section uncertainty are shown in % in Table 5.10. The total systematic uncertainty is calculated as the quadratic sum of all uncertainties. The results are given separately for τ_1 and τ_3. The total uncertainty is of the order of 10-11% for τ_1 and 12% for τ_3. The dominant uncertainties are the tau jet and b-jet identification followed by the MC generator and the jet energy scale.

Table 5.10: *Systematic uncertainties for on the total cross section ($\Delta\sigma/\sigma$) measurement for the fitting method in % for both $\mu + \tau$ and $e + \tau$ channels.*

	$\Delta\sigma/\sigma$ Fitting Method (%)			
	τ_1		τ_3	
	Single p_T bins fit	Combined p_T fit	Single p_T bins fit	Combined p_T fit
μ p_T smearing	0.0	0.0	0.0	0.0
μ Trigger SF	−2.7 / +0.4	−2.7 / +0.4	−2.9 / +0.4	−2.9 / +0.4
μ ID SF	0.0	0.0	0.0	0.0
e p_T smearing	±0.2	±0.2	−0.2 / +0.3	−0.2 / +0.5
e energy scale	±0.3	±0.3	−0.2 / +0.3	−0.2 / +0.4
e Trigger SF	±0.2	±0.2	±0.2	±0.2
e ID SF	±1.2	±1.2	±1.2	±1.2
Jet E scale	−2.8 / +2.3	−2.8 / +2.3	−2.7 / +2.6	−2.9 / +2.6
Jet E res.	±0.1	±0.2	±0.4	±0.3
Jet ID eff.	0.0	0.0	0.0	0.0
$b-$tag	−5.0 / +5.7	−5.0 / +5.7	−5.2 / +5.8	−5.1 / +5.8
ISR/FSR	±4.9	±4.2	±4.3	±4.5
Generator	±2.9	±2.9	±1.04	±1.04
Parton Shower	±4.1	±4.1	±0.6	±0.6
τ ID	±5.0	±4.7	±8.8	±9.8
Total	±11.0	±10.3	±12.1	±12.7

5.7 The $t\bar{t}$ Cross Section

The cross section is derived from the number of measured signal events with the fitting method using the standard definition:

$$\sigma_{t\bar{t}} = \frac{S_{meas}}{A \cdot \epsilon} \cdot \frac{1}{\mathcal{L}}, \tag{5.25}$$

where S_{meas} is the measured signal, A is the signal acceptance, ϵ is the trigger and object reconstruction efficiency, and $\mathcal{L}$ is the integrated luminosity. The branching fraction for $t\bar{t} \to \ell + \tau$ is included in A. In practice, the MC signal model provides by construction most of the terms in Equation 5.25:

$$S_{MC} = A\epsilon\mathcal{L}\sigma_{t\bar{t}(SM)}, \tag{5.26}$$

where S_{MC} is the number of signal events expected by MC and $\sigma_{t\bar{t}(SM)}$ is the standard model $t\bar{t}$ cross section by which the MC was normalized, i.e. $164.57^{+11.45}_{-15.78}$ pb [28], [29]. The actual cross section results are then calculated by

$$\sigma_{t\bar{t}} = \frac{S_{meas}}{S_{MC}} \cdot \sigma_{t\bar{t}(SM)}. \tag{5.27}$$

In section 5.5.2 we obtained the number of signal events in the data sample after applying event selection and b-tagging requirement (S_{Fit}) using the fitting method. S_{Fit} includes the $t\bar{t} \to \ell + \tau$ signal (S_{meas}), and also contributions from non-$t\bar{t}$ τ leptons (mainly $Z \to \tau\tau$ and a small contribution from single top and diboson events) and a non negligible contribution of electrons faking τ leptons from $t\bar{t} \to \ell + e$ and $Z \to ee$. These contributions are shown in table 5.11.

	τ_1		τ_3	
MC Signal	740 ± 6		234 ± 3	
Background τ and e	108 ± 6		28 ± 3	
	Individual fits	Combined fit	Individual fits	Combined fit
Measured τ and e	807 ± 66	802 ± 51	280 ± 52	232 ± 33
Measured τ	699 ± 66	694 ± 51	252 ± 52	204 ± 33
$\frac{S_{meas}}{S_{MC}}$	0.94	0.94	1.07	0.88

Table 5.11: *Measured τ and e objects obtained from the fits. "MC Signal" is the number of expected τ leptons from $t\bar{t} \to \ell + \tau$ (S_{MC}). The "Background τ and e" are the number of τ leptons from non $t\bar{t} \to \ell + \tau$ and e's, while "Measured τ" is the number of τ leptons in $t\bar{t} \to \ell + \tau$ (S_{meas}). $\frac{S_{meas}}{S_{MC}}$ is the ratio of measured to MC signal after subtracting these background τ and e numbers from the measured signal.*

The number of measured signal events can therefore be obtained as:

$$S_{meas} = S_{Fit} - N^{bkg}_{e,\tau}. \tag{5.28}$$

The final cross sections are given separately for τ_1 and τ_3. The systematic uncertainties are taken from table 5.10. The final results for the $t\bar{t} \to \ell + \tau$ cross section measurement using the fitting method are:

$$\sigma(\tau_1) = 155 \pm 15 \text{ (stat.)} \pm 17 \text{ (syst.)} \pm 5 \text{ (lumi.) pb.} \tag{5.29}$$

$$\sigma(\tau_3) = 176 \pm 62 \text{ (stat.)} \pm 21 \text{ (syst.)} \pm 6 \text{ (lumi.) pb.} \tag{5.30}$$

and with the combined fitting technique:

$$\sigma(\tau_1, Comb.) = 154 \pm 12 \text{ (stat.)} \pm 16 \text{ (syst.)} \pm 5 \text{ (lumi.) pb.} \tag{5.31}$$

$$\sigma(\tau_3, Comb.) = 143 \pm 23 \text{ (stat.)} \pm 18 \text{ (syst.)} \pm 5 \text{ (lumi.) pb.} \tag{5.32}$$

The most accurate measurement is obtained with the combined fits technique for τ_1. A combination of the $\sigma(\tau_1)$ and $\sigma(\tau_3)$ measurements have been performed using the Best Linear Unbiased Estimator (BLUE) method [151,152]. The BLUE method has been widely used to combine various observables at the Tevatron and LHC. It adds the input measurements taking into account statistical and systematic uncertainties and their correlations, minimizing the total uncertainty on the combined result. The results or the combination are:

$$\sigma(t\bar{t} \to \ell + \tau) = 156 \pm 15 \text{ (stat.)} \pm 17 \text{ (syst.)} \pm 5 \text{ (lumi.) pb.} \tag{5.33}$$

for the $t\bar{t}$ cross section obtained with the individual p_{T} bins fit technique, and:

$$\sigma(t\bar{t} \to \ell + \tau) = 151 \pm 11 \text{ (stat.)} \pm 15 \text{ (syst.)} \pm 5 \text{ (lumi.) pb.} \tag{5.34}$$

for the $t\bar{t}$ cross section obtained with the combined fits technique.

Combining the $\sigma(\tau_1)$ and $\sigma(\tau_3)$ measurements has an effect of less than 1% in both the statistical and systematic precision.

6
Conclusions

The focus of this thesis has been the measurement of the $t\bar{t}$ production cross section in the lepton plus tau channel ($pp \rightarrow t\bar{t} \rightarrow \ell + \tau$) using 2.05 fb^{-1} of ATLAS data at $\sqrt{s} = 7$ TeV collected during 2011.

Tau leptons play an important role in the search for SM and BSM Higgs bosons. Standard Model neutral Higgs bosons can be produced in association with a top pair, $t\bar{t}H^0$, and decay into two τ leptons: $H^0 \rightarrow \tau\tau$ with a branching fraction of about 10% for low Higgs mass. This decay mode can provide a direct measurement of the coupling of the Higgs boson to fermions, supplying strong evidence that fermions acquire their mass through the Higgs mechanism.

If a charged Higgs boson exists as predicted by the MSSM, and its mass is lower than the top quark mass minus the bottom quark mass, the top quark predominantly decays into a charged Higgs boson and a b-quark: $t \rightarrow H^+b$. A heavy charged Higgs can also be produced associated with a top quark: tH^+. In some scenarios, the charged Higgs would decay predominantly to a tau lepton and a neutrino, producing then an excess in the $\ell + \tau$ channel over the other dilepton channels which, if observed, would constitute experimental evidence of the the existence of a charged Higgs boson.

The $t\bar{t} \rightarrow \ell + \tau$ candidate events are selected based on the event topology of the final state: one high p_{T} electron or muon, 2 high E_{T} jets, at least one of them b-tagged, large $E_{\mathrm{T}}^{\mathrm{miss}}$ and one hadronically decaying τ lepton. The main background after the selection comes from $t\bar{t} \rightarrow \ell$+jet where a jet has been misidentified as a hadronically decaying tau. The only discriminator between this background and the signal is the tau identification. In this analysis we exploit a discriminant variable, BDT_j, obtained from a Boosted Decision Tree multi-variate technique specially trained to separate real taus from other jet types. The signal and background contributions to the signal region are then estimated using a χ^2 fit of the BDT_j distribution of the OS reconstructed taus in the data sample with a signal template derived from MC and background templates derived from data to minimize the systematic uncertainty. Two background templates are used, one for jets produced by the hadronization of light-quarks and another for jets produced

by the hadronization of gluon, which constitute the main source of fake taus for the $t\bar{t} \to \ell + \tau$ signal after the selections. The parameters of the fit are the amount of both backgrounds and the amount of signal and the shapes of the templates are fixed.

The measured $t\bar{t} \to \ell + \tau$ cross section obtained at a center-of-mass energy of 7 TeV for τ_1 and τ_3 are:

$$\sigma(\tau_1) = 154 \pm 12 \text{ (stat.)} \pm 16 \text{ (syst.)} \pm 5 \text{ (lumi.) pb.} \tag{6.1}$$

and:

$$\sigma(\tau_3) = 143 \pm 23 \text{ (stat.)} \pm 18 \text{ (syst.)} \pm 5 \text{ (lumi.) pb.} \tag{6.2}$$

This measurement exceeds the precision of the existing measurements in the same $\ell + \tau$ channel performed by CDF (30%) [65], D$\emptyset$ (25%) [63] and CMS (18%) [66]. Furthermore, it improves the results of previous measurements of the $t\bar{t} \to \ell + \tau$ cross section published at ATLAS [67], which can be seen in Figure 6.1. This former measurement was perform with a fitting method to a OS-SS BDT_j distribution and was cross checked with a Matrix Method, presented in Appendix A. The OS-SS technique allowed the reduction of the background from fake taus to only light-quark jets, thus facilitating the background modelling, since the BDT_j behaviour depends on the type of jet. The background from light-quark jets was derived from data using a control region that is kinematically similar to the signal region but reversing the ≥ 1 b-tag requirement. The contribution of real tau leptons in this region is subtracted based on the predicted MC.

By using the OS data samples for the fits in our analysis we increase the statistics of the fitting templates and reduce the statistical uncertainty of the cross section measurement. We also improve the background description, building a background model which properly reflects the jet composition to correctly estimate the fake τ contamination in the signal region. The systematic uncertainties dominate the measurement and are kept at the same level as previous measurements. The main sources of systematic uncertainties in our analysis are the τ and b-jet identification.

The observed cross section value agree well with the Standard Model expectation of $164.57^{+11.45}_{-15.78}$ pb, within errors, and is consistent with the measured cross-sections using different decay channels by ATLAS [40] [41] [42] [43] and CMS [45] [46] [47], shown in Figure 6.1. Therefore, we cannot make any claims of new physics.

We have, however, developed a technique which has a great potential to achieve even higher precision when applied to high statistic samples, like the 5.08 fb^{-1} of $\sqrt{s} = 7$ TeV data collected by ATLAS in 2011, the 21.3 fb^{-1} of data collected by ATLAS during 2012 with $\sqrt{s} = 8$ TeV or the future 13-14 TeV data that the LHC will produce in Run II.

This method has been applied to the full 5.08 fb^{-1} 2011 dataset to measure the $t\bar{t}$ production cross section in the lepton plus tau channel [153], the top quark decay branching ratios into channels with leptons and jets in $t\bar{t}$ pairs, in order to test their agreement with the Standard Model [154], and the τ identification uncertainty [155]. With more than double the statistics in the sample, this systematic has been reduced by about 40-50%, from 5%(9%) to 2.4%(5.1%) for $\tau_1(\tau_3)$.

The fitting method will improve with higher statistics samples as we will be able to perform a finer binning in p_{T} of the τ candidates without deteriorating the information of each bin. Thus

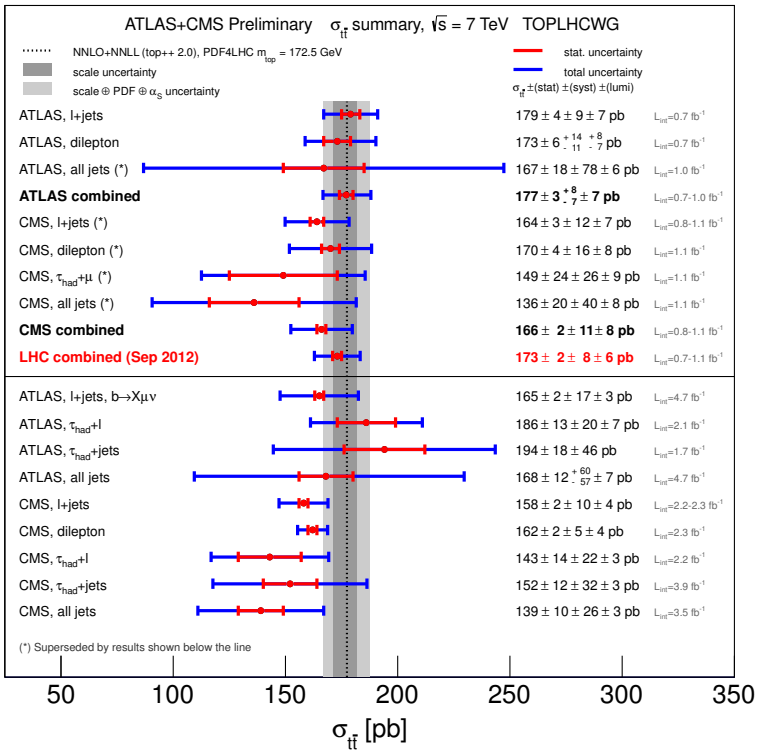


Figure 6.1: *Summary of ATLAS and CMS measurements of the top-pair production cross-section at 7* TeV.

the evolution of the BDT_j shape changes with the p_{T} of the τ candidates will be better tracked and the background modelling will be improved.

7 Resumen

7.1 Motivaciones teóricas

El Modelo Estándar de la física de partículas es una teoría gauge basada en el grupo de simetrías locales SU(3) × SU(2) × SU(1) que describe las partículas fundamentales y sus interacciones [1]. Bajo el Modelo Estándar las partículas fundamentales se describen mediante campos, ψ, que son solución de ecuaciones de Euler-Lagrange, y las interacciones entre dichos campos, fuerte, débil y electromagnética, son mediadas por bosones gauge, cuya masa y otras propiedades se muestran en la Tabla 7.1.

Fuerza	Boson Gauge	Masa [GeV]	Carga [e]	Color	Espín
Fuerte	gluón (g)	0	0	sí	1
Electromagnética	fotón (γ)	0	0	no	1
Débil	Z^0	91.18	0	no	1
	$W^{\pm}$	80.4	± 1	no	1
Gravedad	Gravitón	$< 7 \cdot 10^{-32}$ eV	0	no	2

Table 7.1: *Fuerzas fundamentales y bosones gauge asociados. La gravedad no está incluida en el Modelo Estándar. El Gravitón no ha sido encontrado aún.*

La gravedad es un fuerza de alcance infinito pero muy débil comparada con las otras tres fuerzas. No juega ningún papel en la física de las partículas elementales y no está incluida en el Modelo Estándar.

En el Modelo Estándar los bosones intermediarios de las fuerzas no tienen masa, dado que insertar un término de masa en el lagrangiano rompería la invarianza gauge. Sin embargo, los bosones $W^{\pm}$ y Z^0, mediadores de la interacción débil, tienen masa. Para resolver este problema,

Higgs, Brout y Englert propusieron en 1964 el meanismo de rotura espontánea de simetría o mecanismo de Higgs [9–12]. Las masas de los bosones $W^{\pm}$ y Z^0 son el resultado de su interacción con el campo de Higgs, que puede también proporcionar masa al resto de partículas fundamentales. Un bosón escalar masivo, llamado bosón de Higgs (H), es el mediador de las interacciones con el campo de Higgs. La masa del bosón de Higgs es el único parámetro desconocido de esta teoría.

En Julio de 2012 los experimentos ATLAS y CMS del CERN anunciaron el descubrimiento ($> 5\sigma$) de un nuevo bosón en el rango de masas 125-126 GeV [13] [14], compatible con el bosón de Higgs predicho por el Modelo Estándar.

Las partículas fundamentales son partículas de espín semi-entero llamadas fermiones. Los fermiones susceptibles de experimentar la fuerza fuerte se llaman quarks y aquellos que no interactúan fuertemente se llaman leptones. Los quarks sólo se han observado combinados formando partículas llamadas hadrones, mientras que los leptones pueden existir como partículas libres. En la Tabla 7.2 se muestran las partículas fundamentales con su masa, carga eléctrica y espín.

	Particula	Carga [e]	Color	Masa [MeV]	Espín
Leptones	electrón (e)	-1	no	0.511	1/2
	neutrino electrónico (ν_e)	0	no	< 2 eV	
	muón (μ)	-1	no	105.7	
	neutrino muónico (ν_μ)	0	no	< 2 eV	
	tau (τ)	-1	no	1776.8	
	neutrino tauónico (ν_τ)	0	no	< 2 eV	
Quarks	up (u)	+2/3	sí	2.3	1/2
	down (d)	-1/3	sí	4.8	
	charm (c)	+2/3	sí	$1.275 \cdot 10^3$	
	strange (s)	-1/3	sí	95	
	top (t)	+2/3	sí	$173 \cdot 10^3$	
	bottom (b)	-1/3	sí	$4.18 \cdot 10^3$	

Table 7.2: *Table de partículas elementales.*

7.1.1 Extensiones del Modelo Estándar

El Modelo Estándar ha demostrado ser capaz de describir un amplio abanico de procesos con una gran precisión. Sin embargo, dista de ser una "Teoría del Todo". No proporciona una unificación de las interacciones fuerte y electrodébil y no incluye la gravedad. La materia oscura y las oscilaciones de neutrinos, que han sido establecidas experimentalmente, tampoco tienen cabida dentro del Modelo Estándar.

Aunque la partícula recientemente descubierta es compatible con el bosón de Higgs, no todas sus propiedades han sido medidas con gran precisión, de modo que fenomelogía más allá del Modelo Estándar no puede descartarse aún.

Una de las extensiones más populares del Modelo Estándar es Supersimetría (SUSY), que predice que para cada partícula del Modelo Estándar existe una super-compañera con los mismos

números cuánticos salvo el espín, que se diferencia en $\pm 1/2$. La mínima extensión del Modelo Estándar que incluye supersimetría es el "Minimal Supersymmetric Standard Model" o MSSM. El MSSM predice la existencia de cinco bosones de Higgs, tres neutros (h, H, A) y un par cargado ($H^{\pm}$). Para una revisión más detallada de SUSY y el MSSM se recomienda leer las referencias [19], [20] y [21].

SUSY puede explicar el mecanismo de rotura de la simetría electrodébil, la asimetría entre materia y antimateria y ofrece candidatos viables a materia oscura. Además, SUSY sugiere la unificación de de todas las fuerzas gauge a altas energías. Por todos estos motivos la búsqueda de partículas supersimétricas es una parte importante de los programas de física en el LHC, aunque aún no se ha encontrado ninguna evidencia de su existencia [22].

7.1.2 Física del quark top y relevancia del leptón tau

El quark top es la partícula elemental más masiva y fue descubierta en 1995 por los experimentos CDF [23] y D0 [24] del colisionador de hadrones Tevatron en Fermilab [25]. La cercanía de la masa del top a la escala electrodébil[1] hace posible que el quark top juegue un papel especial en la rotura de simetría electrodébil. Por lo tanto, el quark top es un candidato natural en la búsqueda de fenónomenos de nueva física más allá del Modelo Estándar.

En colisionadores hadrónicos hay dos mecanismos de producción de quarks top: pares top-antitop ($t\bar{t}$), a través de interacciones fuertes gluón-gluón o quark-antiquark, y top individuales ("single top") producidos mediante interacciones débiles. Para la generación de datos simulados y normalización de resultados, en esta tesis se ha aplicado la siguiente sección eficaz de producción $t\bar{t}$, calculada a tercer orden (NNLO) en teoría QCD perturbativa para una energía de centro de masas de $\sqrt{s} = 7$ TeV [28], [29]:

$$\sigma_{t\bar{t}}^{\text{theo}} = 164.57^{+11.45}_{-15.78} \text{ pb}. \tag{7.1}$$

Una vez producido, el quark top se desintegra inmediatamente y sólo se puede detectar a través de los productos de la desintegración. En el Modelo Estándar el quark top se desintegra a un bosón W y un quark b. Los modos de desintegración del bosón W definen los posibles estados finales. Aproximadamente $1/3$ de las veces, el bosón W se desintegra a un leptón (e, μ, τ) y un neutrino ($W \to \ell\nu$). El restante $2/3$ de los casos se desintegra a un par quark/anti-quark ($W \to q\bar{q}$).

Si existe un bosón de Higgs cargado, tal y como predice el MSSM, y su masa es menor que la del quark top, el quark top se desintegra predominantemente a un Higgs cargado y un quark b ($t \to H^+ b$). Bosones de Higgs cargados masivos también pueden producirse en asociación con un quark top (tH^+). Para valores elevados de $\tan\beta$[2] el Higgs se desintegraría predominantemente en el canal $H \to \tau\nu_\tau$. Este proceso produciría un exceso en el canal $\ell + \tau$ sobre los otros canales dileptónicos y la proporción

$$R = \frac{t \to \tau\nu_\tau}{t \to \ell\nu_\ell} \quad (\ell = e, \mu) \tag{7.2}$$

[1] La escala electrodébil está definida por el valor en el vacío del potencial del Higgs Estándar $v = 246$ GeV.

[2] $\tan\beta = \frac{v_u}{v_d}$ se define como el cociente de los valores esperados en el vacío de los potenciales de los dos dobletes de Higgs predichos en el MSSM para el acoplamiento de quarks de tipo "up" y tipo "down".

sería mayor que 1, que es el valor predicho en el Modelo Estándar. Por lo tanto, obtener experimentalmente valores $R > 1$ constitutiría un indicio de la existencia de un bosón de Higgs cargado. La medida de la sección eficaz de producción $t\bar{t} \to \ell + \tau$, que constituye el trabajo central de esta tesis, adquiere pues un papel relevante en búsquedas de nueva física.

Las medidas actuales de la sección eficaz de producción en el canal $t\bar{t} \to \ell + \tau$, tanto en Tevatron como en el LHC, se muestran en la Tabla 7.3.

	$\sqrt{s}$ (TeV)	$\sigma_{t\bar{t}\to\ell+\tau}$ (pb)	$\mathcal{L}_{\text{int}}$ (fb^{-1})	Referencia
CDF	1.96	8.2 ± 2.3 (stat.) $\pm$ 1.2 (syst.) $\pm$ 0.5 (lumi.)	9.0	[65]
DØ	1.96	$7.32\ ^{+1.34}_{-1.24}$ (stat.) $^{+1.49}_{-1.31}$ (syst.) $\pm$ 0.39 (lumi.)	1.2	[63]
ATLAS	7	186 $\pm$ 13 (stat.) $\pm$ 20 (syst.) $\pm$ 7 (lumi.)	2.05	[67]
CMS	7	143 $\pm$ 14 (stat.) $\pm$ 22 (syst.) $\pm$ 3 (lumi.)	2.2	[66]

Table 7.3: *Medida de la sección eficaz de producción de pares top-antitop en el canal leptón + tau en Tevatron y el LHC.*

7.2 El colisionador de hadrones LHC y el detector ATLAS

El gran colisionador de hadrones LHC [68] ha sido construido por el laboratorio europeo para la investigación en física nuclear, CERN [69], junto a la localidad suiza de Ginebra. El LHC es un acelerador y colisionador de partículas circular de 27 km de longitud. Haces de protones generados mediante la ionización de átomos de hidrógeno son conducidos a través de un gran número de aceleradores interconectados que van aumentando progresivamente su energía hasta alcanzar los 450 GeV[3]. En ese momento son inyectados, en direcciones opuestas, en el anillo principal, donde una serie de imanes superconductores se encargan de mantener los haces alineados y de hacerlos colisionar en cuatro puntos de interacción donde se encuentran los detectores. La Figura 7.1 muestra el complejo de aceleradores y los detectores situados en los puntos de colisión.

Las primeras colisiones de protones en el LHC con una energía de centro de masas de 7 TeV tuvieron lugar en 2010. En 2011 se alcanzó una luminosidad instantánea de $\mathcal{L}_{\text{ins}} = 3.65 \cdot 10^{33}\text{cm}^{-2}\text{s}^{-1}$, con un total de 5.46 fb^{-1} proporcionados a los experimentos. En 2012 la energía de centro de masas aumentó hasta 8 TeV y se alcanzó un total de 22.8 fb^{-1}, cuatro veces más que en 2011 y 20 veces mayor que la luminosidad acumulada en Tevatron en 10 años. En la Tabla 7.4 se recogen los parámetros más relevantes del LHC con sus valores nominales y los alcanzados durante el periodo 2010-2012. En 2013 el LHC inició una parada técnica para realizar las actualizaciones necesarias para alcanzar su máximo potencial de funcionamiento.

El LHC alberga 4 grandes detectores localizados en los puntos de colisión: ATLAS, CMS, LHCb y ALICE, y otros dos experimentos menores: LHCf y TOTEM. Los datos analizados y presentados en esta tesis han sido tomados por el detector ATLAS.

[3] Energá mínima a la que el LHC puede mantener los haces estables.

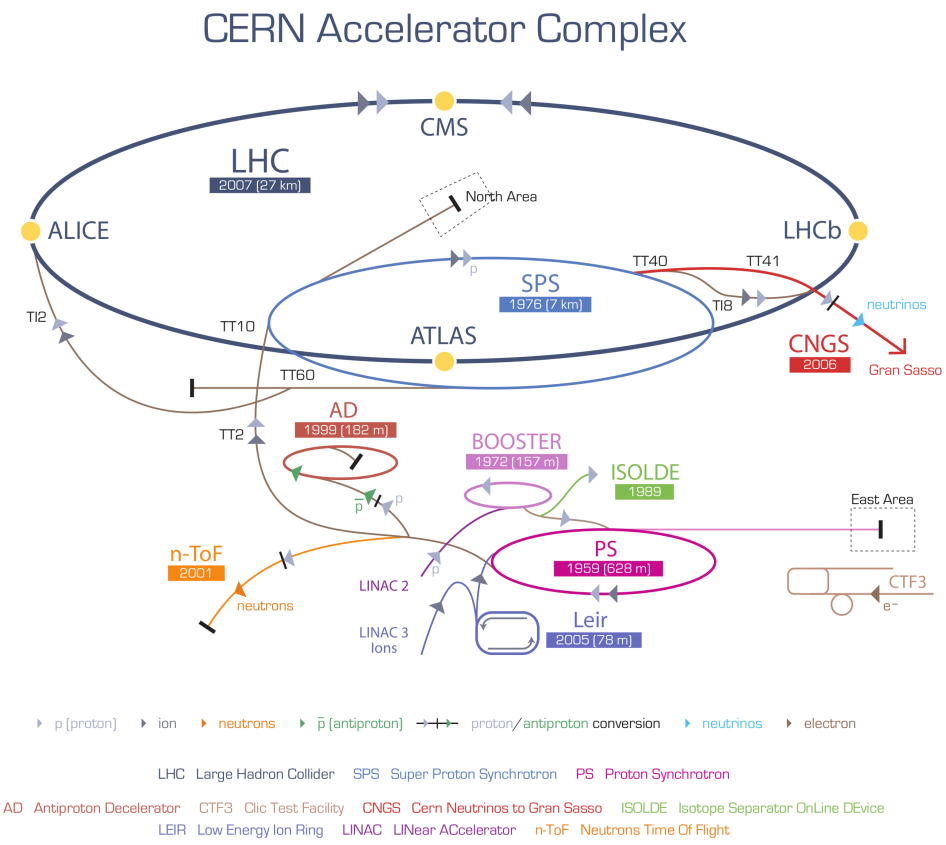


Figure 7.1: *Ilustración del complejo de aceleradores LHC del CERN.*

7.2.1 El detector ATLAS

ATLAS [75] es un detector de propósito general, diseñado para poner a prueba el Modelo Estándar y buscar signos de nueva física. El detector ATLAS tiene forma cilíndrica, mide 42 m de largo por 25 de alto y pesa unas 7000 toneladas. Está formado por varias capas concéntricas de sub-detectores (un detector de trazas, calorímetros electromagnéticos y hadrónicos, y un espectrómetro de muones) como muestra la Figura 7.2. Un solenoide de 2 T cubre el detector de trazas y toroides de hasta 3.5 T revisten el espectrómetro de muones.

El **Detector Interno de trazas** está formado por tres sub-detectores basados en silicio y en tubos de deriva: el detector de píxeles, el detector de trazas semiconductor (SCT) y el detector de trazas de radiación de transición (TRT). Esta configuración está diseñada para medir con precisión el momento, la carga y los vértices primarios y secundarios de las partículas cargadas.

Los **Calorímetros** se encuentran a continuación del selenoide que recubre el detector interno. Hay dos tipos de calorímetro: electromagnético y hadrónico. El calorímetro electromagnético

	Nominal	2010	2011	2012
Energía haces [TeV]	7	3.5	3.5	4
Espaciado paquetes [ns]	25	150	50	50
Número de paquetes	2808	368	1380	1380
Partículas por paquete	$1.15 \cdot 10^{11}$	$1.2 \cdot 10^{11}$	$1.45 \cdot 10^{11}$	$1.7 \cdot 10^{11}$
β^* [m]	0.55	3.5	1.5/1	0.6
ε_n [μm rad]	3.75	2.4	2.4	2.5
γ	7461		3730	
Luminosidad [$cm^{-2}s^{-1}$]	10^{34}	$2.1 \cdot 10^{32}$	$3.7 \cdot 10^{33}$	$7.7 \cdot 10^{33}$
Luminosidad total [fb^{-1}]	-	0.048	5.46	22.8

Table 7.4: *Valores operativos del LHC nominales y alcanzados durante el run I. Se muestra el valor máximo alcanzado cada año [71], [72].*

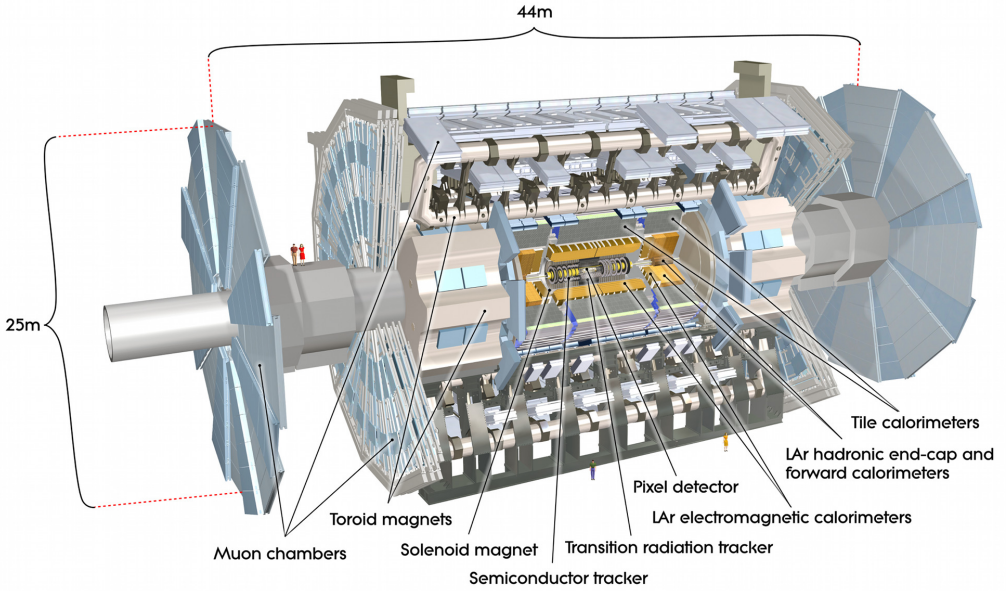


Figure 7.2: *Diseño del detector ATLAS. Los sub-detectores están marcados en el dibujo.*

es un detector de muestreo que utiliza plomo como material absorbente y argón líquido como medio activo. Se encarga de medir la energía depositada por electrones y fotones. El calorímetro hadrónico mide la energía de jets de hadrones, para lo que utiliza una estructura de tejas que intercala plástico centelleador como medio activo y acero como absorbente.

En la parte más exterior de ATLAS se encuentra el **Espectrómetro de Muones**, diseñado para medir el momento de los muones que han escapado del detector interno y los calorímetros.

El detector ATLAS tomó los primeros datos de colisiones a finales de 2009. De un total de 27 fb^{-1} de datos de colisiones producidos por el LHC hasta 2012, ATLAS ha sido capaz de grabar

un total de 23 fb^{-1}, que se corresponde con una eficiencia de toma de datos superior al 90% [82].

Para analizar ingente cantidad de datos producidos la colaboración ATLAS recurre a:

1. Un sistema de "trigger", encargado de discriminar los eventos interesantes desde el punto de vista físico y reducir la cantidad de datos almacenados.

2. Un sistema de computación distribuida basado en tecnologías Grid [87] [88] para analizar y almacenar los datos. La base del sistema Grid es compartir los recursos computacionales entre todos los centros que forman parte del experimento. Los datos son procesados para ofrecer al usuario un formato fácil de analizar que contiene toda la información de los objetos físicos de interés para los diferentes análisis.

7.3 Medida de la sección eficaz de producción de pares top-antitop en el canal de desintegración dileptónico con leptón tau

La sección eficaz de cualquier proceso físico es una medida de la probabilidad de que dicho proceso se produzca en una colisión. Esta probabilidad es proporcional al número de eventos producidos. Para el canal de desintegración $t\bar{t} \to \ell + \tau$ tenemos:

$$\sigma_{t\bar{t}\to\ell+\tau} = \frac{N_{\ell,\tau}}{A \cdot \epsilon} \cdot \frac{1}{\mathcal{L}}, \tag{7.3}$$

dónde $N_{\ell,\tau}$ es el número de eventos $t\bar{t} \to \ell + \tau$ producidos, $\mathcal{L}$ es la luminosidad integrada total, A es la aceptancia del detector a los eventos de señal y ϵ incluye la eficiencia del sistema de trigger, de la reconstrucción de los objetos físicos y de la selección de eventos aplicada para aislar la señal y reducir los fondos.

La clave de esta medida reside en separar la señal que queremos medir:

$$t\bar{t}\to W(\to e/\mu + \nu_{e/\mu})bW(\to\tau + \nu_\tau)b, \tag{7.4}$$

de los procesos de fondo que producen una señal idéntica o similar en el detector:

- Producción de pares top-antitop en los estados finales:
  1. leptón + jets: $t\bar{t}\to W(\to e/\mu + \nu_{e/\mu})bW(\to qq')b$
  2. dileptónico: $t\bar{t}\to W(\to e/\mu + \nu_{e/\mu})bW(\to e/\mu + \nu_{e/\mu})b$
- Producción de quarks top aislados (single top)
- Producción de bosones W y Z con jets asociados
- Producción de pares de bosones: WW, WZ, ZZ

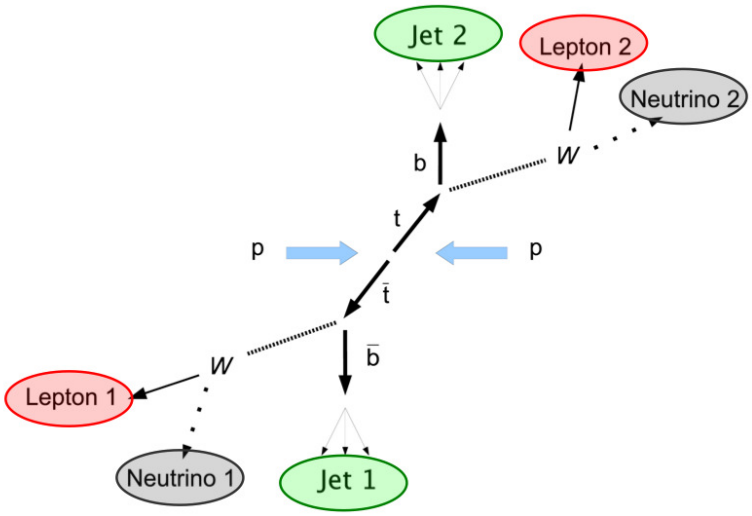


Figure 7.3: *Esquema de la interacción de un par top-antitop desintegrándose a dos leptones. El estado final en el detector se caracteriza por dos b-jets con alta energía transversa E_T, dos leptones aislados (uno de los cuales será un τ que se desintegra hadrónicamente en el canal $t\bar{t} \rightarrow \ell + \tau$) y elevada energía transversa perdida (E_T^{miss}) procedente de los neutrinos.*

7.3.1 Estrategia del análisis

La estrategia empleada en este análisis para la diferenciación de señal y fondos y medida de la sección eficaz es la siguiente:

- Primero realizaremos una selección de eventos dirigida a maximizar la contribución de eventos de señal y minimizar los fondos. Esta selección está detallada en la Sección 7.3.2.

- A continuación determinaremos los procesos de fondo más relevantes para el canal $\ell + \tau$ tras la aplicación de la preselección anterior. Veremos en la Sección 7.3.2 que el fondo dominante es $t\bar{t} \rightarrow \ell$ + jet, donde un jet ha sido reconstruido como un tau.

- Puesto que ambos procesos, $t\bar{t} \rightarrow \ell + \tau$ y $t\bar{t} \rightarrow \ell$ + jet, producen la misma señal en el detector, la única forma efectiva de diferenciarlos es mediante la aplicación de técnicas de identificación de leptones τ. En este análisis utilizaremos una variable discriminante, BDT_j, obtenida mediante técnicas multivariadas de árboles de decisión y entrenada especialmente para separar taus que decaen hadrónicamente de otro tipo de jets [126]. La identificación de leptones tau se ha desarrollado en la Sección 3.3.

- El comportamiento de la variable BDT_j es diferente para cada tipo de jet. Veremos en la sección 7.3.3 que los tipos de jet dominantes en el fondo son quarks ligeros y gluones y obtendremos las distribuciones de BDT_j para ambos a partir de la muestra de datos. La distribución BDT_j para taus reales la obtendremos de Monte Carlo.

- Finalmente, haremos un ajuste de la distribución BDT_j de la muestra de datos con las distribuciones de BDT_j obtenidas para los fondos y la señal:

$$\text{Data} = k_s \cdot \text{signal} + k_g \cdot \text{Gluons} + k_q \cdot \text{light-quark}. \tag{7.5}$$

  El resultado del ajuste nos proporcionará el número de eventos $t\bar{t} \to \ell + \tau$ en la muestra de datos, con el que calcularemos la sección eficaz de producción.

Esta técnica de ajustes pretende mejorar los resultados ya publicados por ATLAS para la sección eficaz de producción $t\bar{t} \to \ell + \tau$ [67] [135], mejorando la descripción de los fondos y reduciendo el error estadístico, a la vez que se mantienen las incertidumbres sistemáticas.

Los datos analizados corresponden a colisiones protón-protón, con una energía de centro de masas $\sqrt{s} = 7$ TeV, registradas por ATLAS entre Marzo y Agosto de 2011. La luminosidad integrada total de los datos es 2.05 fb^{-1}. Se han utilizado muestras de datos simulados o Monte Carlo para optimizar los procesos de selección, calcular la aceptancia de la señal y evaluar las contribuciones de los fondos de single top, dibosones, $W+$ jet y $Z \to \tau^+\tau^-$. Las especificaciones de la generación y el procesado de las muestras de datos y Monte Carlo pueden consultarse en el Capítulo 3.

7.3.2 Selección de eventos

La selección de eventos para el análisis de los datos tomados en 2011 ha sido optimizada para tratar de aislar la topología concreta de nuestros eventos de señal, tal y como se puede observar en la Figura 7.3: b-jets de alta E_{T}, un leptón de alto p_{T}, un τ que decae hadrónicamente y elevada $E_{\mathrm{T}}^{\mathrm{miss}}$. Los detalles de la selección de objetos (leptones, jets, jets procedentes de quarks b, energía transversa perdida y taus reconstruidos) pueden consultarse en el Capítulo 3.

La selección de eventos $t\bar{t} \to \ell + \tau$ comienza online aceptando únicamente eventos seleccionados por el trigger de leptones: eventos que contienen al menos un muón con $p_{\mathrm{T}} > 18$ GeV o un electrón con $p_{\mathrm{T}} > 20$ GeV.

La selección offline ha sido optimizada usando muestras de Monte Carlo:

- Para asegurar la calidad de los objetos reconstruidos los datos han de verificar ciertos criterios de calidad codificados en una serie de etiquetas que reflejan el estado de cada detector, el sistema de trigger y cada objeto físico reconstruido (jets, electrones, muones, etc). Estas etiquetas, que reflejan el estado del detector ATLAS, se guardan para cada bloque de luminosidad [130]. Los bloques de luminosidad que pasan los criterios de calidad se almacenan en una lista (Good Run List o GRL) y los eventos no incluidos en dicha lista son descartados.

- Para asegurar que los eventos proceden de colisiones de protones se requiere que los eventos posean un vértice primario con al menos 5 trazas, cada una con $p_{\mathrm{T}} > 400$ MeV. Por la misma razón rechazamos aquellos eventos con al menos un jet con $p_{\mathrm{T}} > 20$ GeV que no pasa los los criterios de calidad diseñados para rechazar jets reconstruidos en zonas de ruido en el calorimetro o procedentes de otros eventos debido al pileup [130].

- Un único lepton aislado en el evento. Un muón aislado para el canal $\mu + \tau$ y un electrón aislado para el canal $e + \tau$. Este requisito reduce abundantemente el fondo de multi-jets.

- Al menos un τ reconstruido (definido en la Sección 3.2).
- Al menos 2 jets con $p_{\mathrm{T}} > 25$ GeV que no sobrelapen con un τ reconstruido, es decir, $\Delta R(\tau, jet) > 0.4$. En caso de haber 2 τ reconstruidos se exige al menos 1 jet que no sobrelape con ninguno de los dos y ambos taus se conservan hasta la identificación.
- $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV para reducir el fondo de multi-jets QCD. Debido a la elevada sección eficaz de producción de multi-jets, el fondo procedente de jets misidentificados como leptones aislados no es despreciable, pero se puede reducir requiriendo una elevada energía transversa perdida, debida a la presencia de neutrinos energéticos en nuestra señal que no aparecen en los procesos de multi-jets.
- H_{T}[4] > 200 GeV, para reducir el fondo procedente de W + jets.
- Al menos un jet identificado como procedente de la desintegración de un quark b (≥ 1 b-tag) aplicando un corte ≥ 0.35 sobre el peso del jet proporcionado por el algoritmo CombNN, que tiene una eficiencia del 70%. Por simplicidad de lenguaje, nos referiremos a la identificación de estos jet por su nombre en inglés: b-tagging.

Esta preselección define lo que llamaremos la región de señal ≥ 1 b-tag.

Las Tablas 7.5 a 7.8 muestran el número de eventos observado y esperado tras cada paso de la selección. Los fondos dominantes antes de la aplicación del b-tagging son $t\bar{t} \rightarrow \ell$ + jet, con un 70% de los eventos y W + jets, con un 20% de los eventos. La aplicación del b-tagging prácticamente erradica el fondo W + jets, de modo que $t\bar{t} \rightarrow \ell$ + jet es el fondo dominante. En este punto la única forma de diferenciar la señal del fondo es mediante la identificación de leptones τ.

[4] suma del p_{T} del leptón, el τ reconstruido y los jets y $E_{\mathrm{T}}^{\mathrm{miss}}$

Table 7.5: *Selección de eventos en el canal $\mu + \tau$ para taus 1-prong. $t\bar{t}(\ell\ell')$ son eventos $t\bar{t}$ dileptónicos con un único leptón reconstruido como un muón y un leptón o un jet reconstruidos como un τ. Las últimas dos filas muestran los números individualmente para eventos OS y eventos SS que pasan el corte final o b-tagging. Los errores corresponden a la incertidumbre estadística.*

Cut	$t\bar{t}(\mu,\tau)$	$t\bar{t}(\ell+\text{jets})$	$t\bar{t}(\ell\ell')$	Z+jets	W+jets	Single top	Diboson	Total	Data
Isolated μ	3972 ± 14	25938 ± 37	3621 ± 14	739203 ± 698	8398984 ± 9931	11663 ± 57	12042 ± 60	9195423 ± 9956	15339738
≥ 1 τ candidate	784 ± 6	3346 ± 13	169 ± 3	20537 ± 109	61832 ± 395	721 ± 13	1198 ± 19	88586 ± 411	258721
$N_{jet} \geq 2$	641 ± 6	3189 ± 13	138 ± 3	2193 ± 35	12396 ± 112	379 ± 8	287 ± 9	19224 ± 119	28255
$E_T^{miss} > 30$ GeV	559 ± 5	2561 ± 12	124 ± 2	958 ± 23	8929 ± 92	303 ± 7	203 ± 8	13637 ± 96	15700
$H_T + E_T^{miss} > 200$ GeV	550 ± 5	2546 ± 12	123 ± 2	764 ± 21	7614 ± 80	290 ± 7	177 ± 7	12065 ± 85	12997
≥ 1 bjet	472 ± 5	2105 ± 11	98 ± 2	51 ± 5	712 ± 31	204 ± 6	16 ± 2	3658 ± 35	3809
≥ 1 bjet (OS)	431 ± 5	1506 ± 9	54 ± 2	33 ± 4	440 ± 24	147 ± 5	10 ± 2	2621 ± 27	2475
≥ 1 bjet (SS)	41 ± 1	599 ± 6	44 ± 1	17 ± 3	272 ± 21	57 ± 3	6 ± 1	1037 ± 22	1334

Table 7.6: *Selección de eventos en el canal $\mu + \tau$ para taus 3-prong. $t\bar{t}(\ell\ell')$ son eventos $t\bar{t}$ dileptónicos con un único leptón reconstruido como un muón y un leptón o un jet reconstruidos como un τ. Las últimas dos filas muestran los números individualmente para eventos OS y eventos SS que pasan el corte final o b-tagging. Los errores corresponden a la incertidumbre estadística.*

Cut	$t\bar{t}(\mu,\tau)$	$t\bar{t}(\ell+\text{jets})$	$t\bar{t}(\ell\ell')$	Z+jets	W+jets	Single top	Diboson	Total	Data
Isolated μ	3972 ± 14	25938 ± 37	3621 ± 14	739203 ± 698	8398984 ± 9931	11663 ± 57	12042 ± 60	9195423 ± 9956	15339738
≥ 1 τ candidate	662 ± 6	8210 ± 21	522 ± 5	29086 ± 132	156968 ± 588	1867 ± 21	2246 ± 26	199562 ± 604	700853
$N_{jet} \geq 2$	508 ± 5	7733 ± 20	413 ± 5	3820 ± 45	31896 ± 178	892 ± 13	569 ± 13	45831 ± 186	71154
$E_T^{miss} > 30$ GeV	446 ± 5	6203 ± 18	371 ± 4	1525 ± 29	23116 ± 148	705 ± 11	398 ± 11	32765 ± 153	39118
$H_T + E_T^{miss} > 200$ GeV	440 ± 5	6168 ± 18	366 ± 4	1278 ± 26	20192 ± 131	680 ± 11	347 ± 10	29473 ± 136	33007
≥ 1 bjet	370 ± 4	5141 ± 16	295 ± 4	98 ± 7	1973 ± 50	478 ± 9	42 ± 3	8397 ± 54	9410
≥ 1 bjet (OS)	240 ± 4	3350 ± 13	147 ± 3	46 ± 5	1120 ± 37	309 ± 7	28 ± 3	5240 ± 40	5712
≥ 1 bjet (SS)	129 ± 3	1791 ± 10	149 ± 3	51 ± 5	854 ± 34	169 ± 6	14 ± 2	3157 ± 37	3698

Table 7.7: *Selección de eventos en el canal $e + \tau$ para taus 1-prong. $t\bar{t}(\ell\ell')$ son eventos $t\bar{t}$ dileptónicos con un único leptón reconstruido como un electrón y un leptón o un jet reconstruidos como un τ. Las últimas dos filas muestran los números individualmente para eventos OS y eventos SS que pasan el corte final o b-tagging. Los errores corresponden a la incertidumbre estaística.*

Cut	$t\bar{t}(\mu,\tau)$	$t\bar{t}(\ell + \text{jets})$	$t\bar{t}(\ell\ell')$	Z+jets	W+jets	Single top	Diboson	Total	Data
Isolated e	3604 ± 14	23501 ± 35	3230 ± 13	770763 ± 709	6878437 ± 9081	9792 ± 51	9907 ± 54	7699235 ± 9109	10960533
≥ 1 τ candidate	695 ± 6	3018 ± 13	151 ± 3	21889 ± 114	93486 ± 835	623 ± 11	944 ± 17	120806 ± 843	228810
$N_{jet} \geq 2$	572 ± 5	2874 ± 12	124 ± 2	3952 ± 47	10604 ± 122	342 ± 8	268 ± 8	18737 ± 132	29306
$E_T^{miss} > 30$ GeV	495 ± 5	2249 ± 11	111 ± 2	1365 ± 28	7198 ± 100	254 ± 7	165 ± 7	11837 ± 105	14033
$H_T + E_T^{miss} > 200$ GeV	490 ± 5	2238 ± 11	110 ± 2	1199 ± 26	6128 ± 77	246 ± 6	147 ± 6	10558 ± 83	11926
≥ 1 bjet	416 ± 5	1850 ± 10	89 ± 2	64 ± 6	540 ± 26	171 ± 5	14 ± 2	3143 ± 29	3373
≥ 1 bjet (OS)	377 ± 4	1307 ± 8	51 ± 2	46 ± 5	316 ± 21	122 ± 4	9 ± 1	2227 ± 24	2269
≥ 1 bjet (SS)	40 ± 1	543 ± 5	38 ± 1	17 ± 3	224 ± 15	49 ± 3	5 ± 1	916 ± 17	1104

Table 7.8: *Selección de eventos en el canal $e + \tau$ para taus 3-prong. $t\bar{t}(\ell\ell')$ son eventos $t\bar{t}$ dileptónicos con un único leptón reconstruido como un electrón y un leptón o un jet reconstruidos como un τ. Las últimas dos filas muestran los números individualmente para eventos OS y eventos SS que pasan el corte final o b-tagging. Los errores corresponden a la incertidumbre estaística.*

Cut	$t\bar{t}(\mu,\tau)$	$t\bar{t}(\ell + \text{jets})$	$t\bar{t}(\ell\ell')$	Z+jets	W+jets	Single top	Diboson	Total	Data
Isolated e	3604 ± 14	23501 ± 35	3230 ± 13	770763 ± 709	6878437 ± 9081	9792 ± 51	9907 ± 54	7699235 ± 9109	10960533
≥ 1 τ candidate	589 ± 6	7437 ± 20	471 ± 5	32595 ± 140	237909 ± 1311	1597 ± 19	1928 ± 24	282526 ± 1319	630596
$N_{jet} \geq 2$	453 ± 5	7015 ± 19	371 ± 4	8261 ± 69	28143 ± 217	836 ± 12	536 ± 12	45613 ± 229	73435
$E_T^{miss} > 30$ GeV	395 ± 5	5502 ± 17	331 ± 4	2612 ± 38	19262 ± 175	635 ± 11	327 ± 10	29065 ± 180	34471
$H_T + E_T^{miss} > 200$ GeV	392 ± 5	5476 ± 17	329 ± 4	2409 ± 37	16984 ± 155	619 ± 11	294 ± 9	26502 ± 161	29930
≥ 1 bjet	329 ± 4	4544 ± 16	263 ± 4	148 ± 9	1707 ± 50	435 ± 9	32 ± 3	7458 ± 54	8218
≥ 1 bjet (OS)	214 ± 3	2955 ± 13	132 ± 3	83 ± 7	965 ± 36	272 ± 7	17 ± 2	4639 ± 40	5030
≥ 1 bjet (SS)	115 ± 2	1589 ± 9	131 ± 3	66 ± 6	742 ± 34	163 ± 5	14 ± 2	2820 ± 36	3188

7.3.3 Determinación de los fondos mediante técnicas derivadas de datos

El fondo dominante tras aplicar la selección de eventos descrita en 7.3.2 es $t\bar{t} \to \ell$ + jet donde un jet ha sido reconstruido como un tau. Si analizamos el origen de los tau reconstruidos usando las muestras simuladas de Monte Carlo, observamos, tal y como se ve en la Figura 7.4, que tenemos contribuciones de τ reales, quarks ligeros, gluones y una pequeña contribución de b-jets. Si los separamos según el signo relativo de la carga del tau y el leptón en OS (cargas de signo opuesto, representados en el eje positivo) y SS (cargas de igual signo, representados en el eje negativo), los leptones τ reales contribuirán únicamente en los casos OS, mientras que el resto de jets contribuyen en ambos casos.

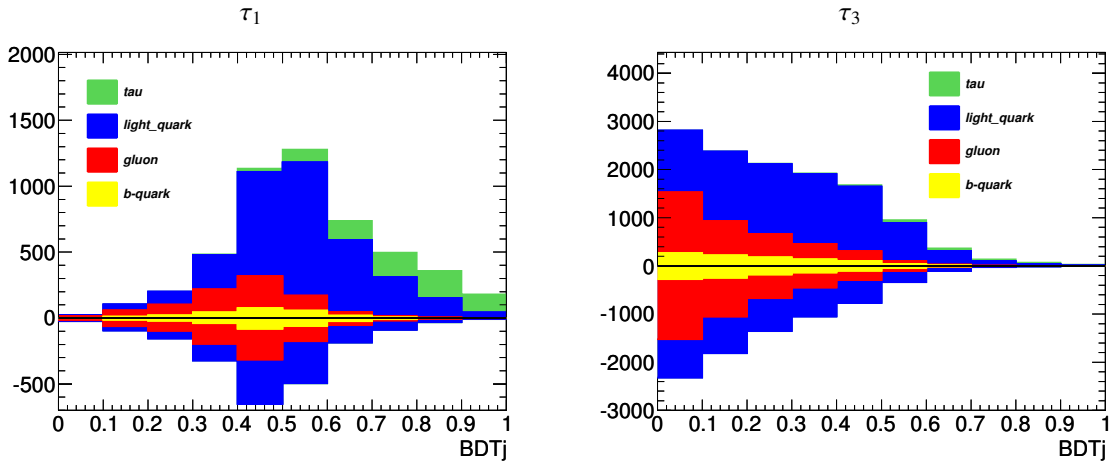


Figure 7.4: *Distribución BDT_j de los τ reconstruidos en función del tipo real de jet. En el eje positivo se muestran los τ con carga de signo opuesto al leptón (OS) y en el eje negativo los τ con carga del mismo signo (SS). Los τ reales únicamente contribuyen a la distribución OS, mientras que los falsos tau contribuyen tanto a las distribuciones OS como a las SS.*

La contribución de leptones tau reconstruidos y originados por gluones y b-jets es la misma en los casos OS y SS debido a la simetría de carga. Las medidas de la sección eficaz $t\bar{t} \to \ell + \tau$ publicadas por ATLAS [67] [135] explotaron esta simetría de carga para substraer las distribuciones BDT_j de la muestra SS de las distribuciones BDT_j de la muestra OS, eliminando las contribuciones de gluones y b-jets, sin perder ningún tau real. Así, los quarks ligeros se convierten el único tipo de jet que contribuye al fondo.

En el presente análisis trabajamos con las distribuciones BDT_j OS, aumentando la estadística de la muestra y reduciendo el error estadístico de la medida. Las distribuciones de BDT_j para gluones y quarks ligeros se modelizan a partir de la muestra de datos para evitar la incertidumbre sistématica asociada a la composición de jets en los modelos de MC (ver Apéndice B). De este modo mejoramos la modelización del fondo para el canal $t\bar{t} \to \ell + \tau$.

Para medir el número de taus reales en la muestra de datos ajustamos la distribución BDT_j de datos usando la técnica de minimización χ^2:

$$\mathrm{OS_{Data}} = k_s \cdot \mathrm{signal} + k_g \cdot \mathrm{Gluons} + k_q \cdot \mathrm{OS_q}, \tag{7.6}$$

donde $\mathrm{OS_{Data}}$, signal and $\mathrm{OS_q}$ son las distribuciones BDT_j OS para datos, señal y quarks ligeros, respectivamente, y Gluons es la distribución BDT_j de los gluones.

Los ajustes se realizarán por separado para τ_1 y τ_3 puesto que las distribuciones de BDT_j tienen formas muy diferentes.

Fondo de quarks ligeros

A partir de la muestra de datos podemos escoger eventos del tipo W + 1 jet y W + 2 jets aplicando la siguiente selección:

- Un leptón aislado (electrón o muón)
- Un único τ reconstruido (que no se corresponda con un b-jet identificado)
- $E_{\mathrm{T}}^{\mathrm{miss}} \geq 30$ GeV para reducir el fondo de multi-jets QCD
- $40 \leq M_{\mathrm{T}}(E_{\mathrm{T}}^{\mathrm{miss}}, \ell) \leq 90$ GeV para rechazar eventos procedentes de sucesos Z + jets.
- Si hay un único jet en el evento y es un τ reconstruido, el evento se clasifica como W + 1 jet. Si hay dos jets y el de menor p_{T} es un τ reconstruido el evento se considera W + 2 jets.

Las muestras W + jets están constituidas por diferentes proporciones de gluones y quarks ligeros. La muestra W + 1 jet OS es rica en quarks ligeros y con ella modelizamos el fondo de quarks ligeros. La muestra W + 2 jets presenta un mayor contenido de gluones y junto con la muestra W + 1 jet nos permitirá obtener la distribución BDT_j para el fondo de gluones.

Fondo de gluones

Puesto que la distribución BDT_j de los gluones presenta simetría de carga, tiene la misma forma y contribuye con el mismo número de eventos a las muestras OS y SS tanto para W + 1 jet como para W + 2 jets:

$$\mathrm{OS1} = a_1 \cdot \mathrm{OS_q} + b_1 \cdot \mathrm{Gluons}, \tag{7.7}$$

$$\mathrm{SS1} = c_1 \cdot \mathrm{SS_q} + b_1 \cdot \mathrm{Gluons}, \tag{7.8}$$

$$\mathrm{OS2} = a_2 \cdot \mathrm{OS_q} + b_2 \cdot \mathrm{Gluons}, \tag{7.9}$$

$$\mathrm{SS2} = c_2 \cdot \mathrm{SS_q} + b_2 \cdot \mathrm{Gluons}, \tag{7.10}$$

donde OS1 (SS1) es la muestra W + 1 jet OS (SS) y OS2 (SS2) es la muestra W + 2 jets OS (SS). $\mathrm{OS_q}$ y $\mathrm{SS_q}$ representan las contribuciones de quarks ligeros.

A partir de estas ecuaciones podemos obtener la distribución Gluons de forma independiente para OS y SS:

$$K \cdot \mathrm{Gluons} = (N \cdot \mathrm{OS2} - \mathrm{OS1}), \tag{7.11}$$

$$K \cdot \text{Gluons} = (N \cdot \text{SS2} - \text{SS1}), \tag{7.12}$$

donde N es el número de eventos en la distribución OS1-SS1 dividido por el número de eventos en la distribución OS2-SS2 y $K = N \cdot b_2 - b_1$ es una constante desconocida. Puesto que ambas distribuciones son idénticas podemos sumarlas y reducir la incertidumbre estadística del fondo de gluones.

7.3.4 Determinación de la señal mediante técnicas de ajuste

El ajuste de los datos se realiza aplicando la técnica de minimización χ^2. El método χ^2 mide de la desviación total de los valores observados y_i respecto de los predichos por el modelo f_i. La descripción óptima de los datos es aquella que minimiza dichas desviaciones:

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - f_i)^2}{\sigma_{y_i}^2 + \sigma_{f_i}^2}, \tag{7.13}$$

donde N es el número de medidas (número de bines en las distribuciones BDT_j), f_i es el valor de la combinación de señal y fondos en cada bin:

$$f_i = k_s \cdot \text{signal} + k_g \cdot \text{Gluons} + k_q \cdot \text{OS}_\text{q}, \tag{7.14}$$

y σ_{y_i}, σ_{f_i} son las incertidumbres de los datos y f en el bin i, respectivamente.

Los valores óptimos de los parámetros de normalización k_g, k_q and k_s se hallan de forma iterativa, minimizando χ^2. Se trata de un ajuste con 2 parámetros libres, ya que el número total de eventos de señal y fondo está constreñido por el número de eventos en la muestra de datos. El proceso de minimización se ha realizado con el programa MINUIT, ampliamente utilizado en el campo de la física de partículas [139] e incluido en la infraestructura de ROOT.

La forma de las distribuciones BDT_j varía con el p_T de los τ reconstruidos, como se puede observar en la Figura 5.8. Por lo tanto, la estrategia que seguimos es construir las distribuciones de BDT_j para datos, señal y fondos en distintos rangos de p_T y hacer los ajustes individualmente para cada uno. Debido a la estadística limitada de la muestra de 2 fb^{-1} empleamos únicamente 2 bines de p_T: $20 < p_\text{T} < 35$ y $35 < p_\text{T} < 100$ GeV.

La comparación de las distribuciones BDT_j en los canales $e + \tau$ y $\mu + \tau$ muestra un buen acuerdo entre ellas (Figura 5.9), tal y como se esperaba ya que la forma de la distribución BDT_j de los τ reconstruidos no depende del canal. Por lo tanto, construimos las distribuciones BDT_j conjuntamente para los canales $e + \tau$ y $\mu + \tau$, con la consecuente ganancia de estadística, y los ajustes se realizan simultaneamente en los dos canales.

Comprobación de la técnica con muestras simuladas de Monte Carlo

Antes de aplicar el método de ajustes a los datos reales tomados por el detector ATLAS, comprobamos con muestras de MC que la señal medida coincide con la cantidad de señal de la muestra. Lo que obtenemos es una desviación del 9% para τ_1 y 14% para τ_3, respectivamente, como se puede observar en la Tabla 7.9. El origen de estas desviaciones es la modelización del fondo. Haciendo un ajuste del fondo esperado de MC con los fondos de gluones y quark ligeros obtenidos con las selecciones definidas en 7.3.3 obtenemos las correcciones que habría que aplicar para que ambos fondos coincidieran. Estas correcciones se pueden ver en la Figura

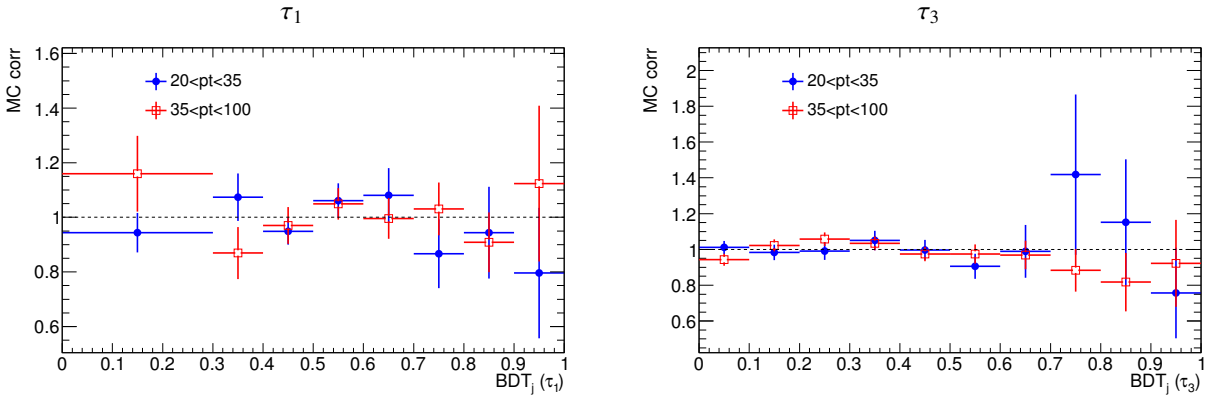


Figure 7.5: *Factores de corrección obtenidos al ajustar la distribución BDT$_j$ del fondo esperado de MC con los fondos obtenidos a partir de datos en el apartado 7.3.3.*

7.5. Tras aplicar las correcciones de MC a los fondos y rehacer los ajustes, no se aprecia ninguna desviación en la señal medida. Esta mejoría justifica la decisión de aplicar los factores de corrección de MC a los fondos derivados de la muestra de datos para extraer la señal en la muestra de 2 fb^{-1}.

Valores esperados	$20 < p_T < 35$ GeV	$35 < p_T < 100$ GeV	Total
τ_1	409 ± 6	440 ± 6	849 ± 8
τ_3	112 ± 2	152 ± 4	264 ± 4
Ajustes sin correcciones			
τ_1	357 ± 50	429 ± 37	786 ± 62
τ_3	107 ± 30	111 ± 28	218 ± 41
Ajustes con correcciones			
τ_1	409 ± 52	440 ± 44	849 ± 68
τ_3	112 ± 32	152 ± 32	264 ± 45

Table 7.9: *Valores esperados de señal en la muestras MC y el obtenido con los ajustes.*

Aplicación de la técnica de ajuste a datos reales

El resultado de los ajustes realizados con la muestra de datos de 2.05 fb^{-1} y los fondos corregidos por MC para extraer la señal $t\bar{t} \to \ell + \tau$ está sintetizado en la Tabla 7.10. Los ajustes en ambos bines de p_T, para τ_1 y τ_3, se muestran en la Figura 7.6: señal en azul, gluones en naranja y W + 1 jet en rosa. Los puntos representan los datos y las líneas punteadas la incertidumbre estadística de los ajustes.

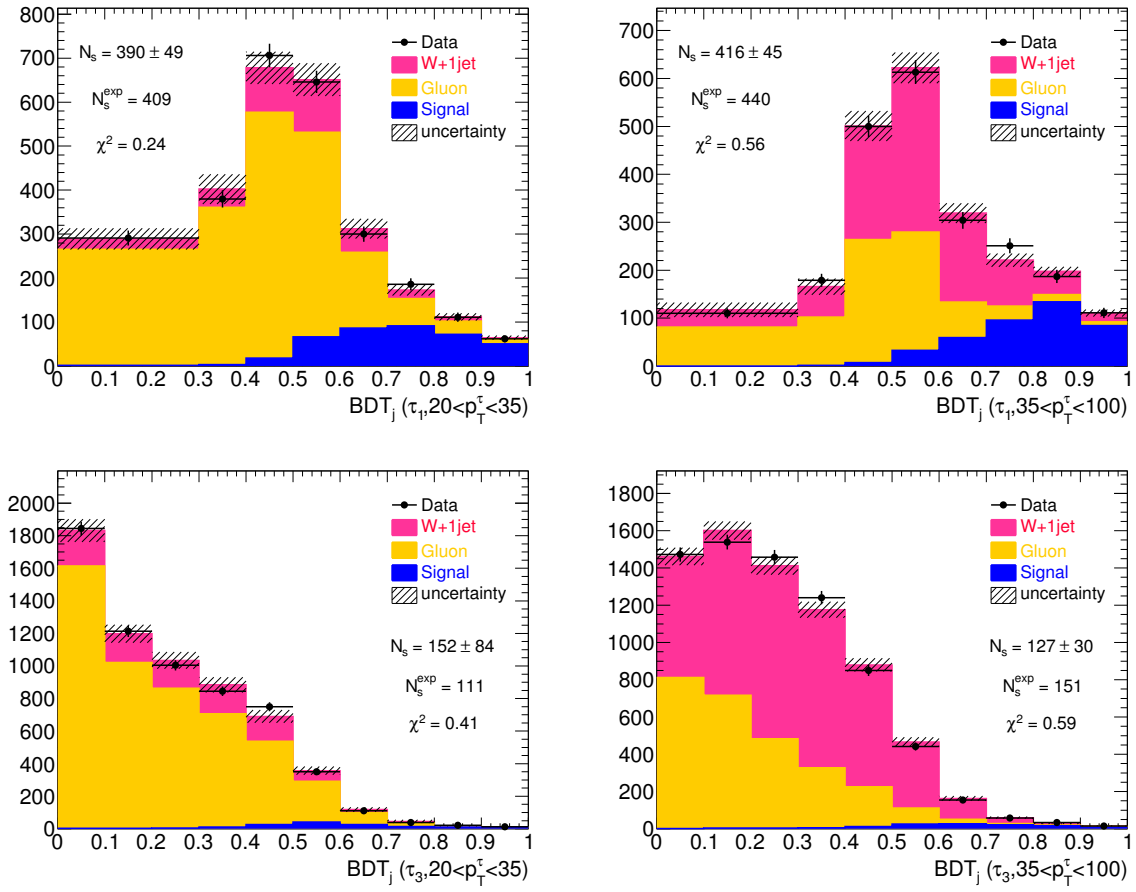


Figure 7.6: *Ajuste de las distribuciones BDT_j en la muestra de datos para τ_1 y τ_3. La columna de la izquierda muestra los leptones τ con p_T entre 20 y* 35 GeV. *La columna de la derecha muestra los leptones τ con p_T entre 35 y* 100 GeV.

Técnica de ajustes combinados

Un ajuste combinado en las dos regiones de p_T ($20 < p_T < 35$ GeV and $35 < p_T < 100$ GeV) nos ha permitido reducir considerablemente el error estadístico en la medida de la sección eficaz. En lugar de realizar dos ajustes independientes, con un total de 4 parámetros libres, hacemos un único ajuste con 3 parámetros libres, tras imponer la restricción de que la suma del número total de eventos de señal y fondo debe ser igual al número total de eventos observados en cada región de p_T. En este tipo de ajuste hay un único parámetro libre para la normalización de las distribuciones de señal, y en cada región de p_T se pesa con la aceptancia relativa de la señal, predicha por MC.

La minimización χ^2 en este caso es como la descrita por la ecuación 7.13, pero la función f para cada bin de p_T viene dada por:

$$f_j = k_s \cdot \frac{N^j_{\text{signal}}}{N_{\text{signal}}} \cdot \text{signal} + k_{gj} \cdot \text{Gluons} + k_{qj} \cdot \text{OS}_\text{q}, \tag{7.15}$$

donde N_{signal} es el número total de eventos de señal en todo el rango de p_T y N^j_{signal} es el número total de eventos de señal en cada bin de p_T.

El resultado de los fits combinados se muestra en la Figura 7.7 y están resumidos en la Tabla 7.10. La señal extraída se encuentra dentro de la incertidumbre de los ajustes individuales y la incertidumbre estadística se ha reducido un 23% en el caso 1-prong y un 64% en el caso multi-prong.

	$20 < p_\text{T} < 35$ GeV	$35 < p_\text{T} < 100$ GeV	Merged p_T bins	Combined p_T Fits	MC
τ_1	391 ± 49 (0.24)	417 ± 45 (0.56)	807 ± 66	802 ± 50 (0.39)	848 ± 8
τ_3	153 ± 84 (0.41)	128 ± 39 (0.59)	280 ± 92	232 ± 32 (0.68)	262 ± 4

Table 7.10: *Resultados del ajuste de las distribuciones BDT_j en la muestra de datos $\ell + \tau$. La segunda y tercera columnas muestran el resultado de los ajustes en los dos bines de p_T. La cuarta columna muestra la señal total extraida. Los números en paréntesis son el valor mínimo de χ^2 obtenido. Los resultados del ajuste combinado se muestran en la quinta columna. El número de eventos esperado para una sección eficaz teórica de 164 pb (MC) se muestra en la sexta columna.*

7.3.5 Discusión de los errores sistemáticos

La medida de la sección eficaz se ve afectada por errores sistamáticos debidos a incertidumbres relacionadas con el rendimiento del detector, la fiabilidad de las simulaciones de Monte Carlo y la incertidumbre del propio método de análisis.

Incertidumbres relacionadas con la simlacón de Monte Carlo

Las principales fuentes de incertidumbre relacionadas con MC son la elección del generador [140–142], la función de distribución de partones (PDF) empleada [143–145] y la modelización de la radiación en los estados inicial (ISR) y final (FSR) [95]. La incertidumbre dominante en esta categoría es la modelización ISR/FSR, con un 4.5%, aproximadamente, de la incertidumbre total en la medida de la sección eficaz, como se puede observar en la Tabla 7.11.

Incertidumbres relacionadas con el rendimiento del detector

El rendimiento del detector limita la eficiencia de reconstrucción e identificación de las partículas y la resolución de la energía y el momento de las mismas.

Las incertidumbres en la eficiencia de reconstrucción e identificación de electrones y muones y su resolución energética se han calculado utilizando muestras simuladas $Z/\gamma \rightarrow \ell^+\ell^-$ donde $\ell = e$ o μ [147], [148]. El efecto sobre la sección eficaz es prácticamente despreciable, al igual que las incertidumbres de identificación de jets y E^{miss}_T [117].

La incertidumbre en la medida de la escala energética de los jets, JES, tiene en cuenta efectos debidos a zonas muertas en el detector y otros defectos, así como la incertidumbre inherente a

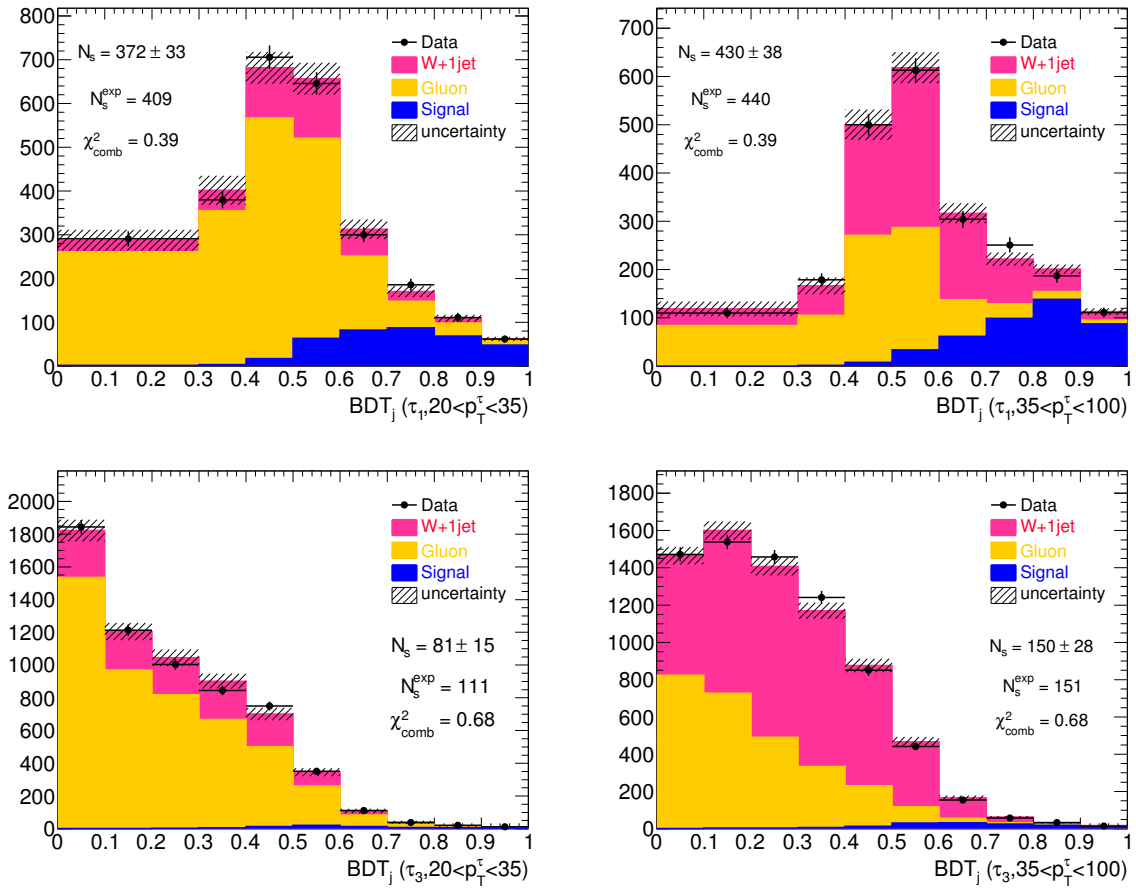


Figure 7.7: *Ajuste de las distribuciones BDT_j en la muestra de datos usando la técnica de ajustes combinados para τ_1 y τ_3. La columna de la izquierda muestra los leptones τ con p_T entre 20 y* 35 GeV. *La columna de la derecha muestra los leptones τ con p_T entre 35 y* 100 GeV.

los algoritmos de reconstrucción [149, 150]. Varía en el rango 4-8% en función de los valores de p_T y η del jet y produce un 3% de la incertidumbre total de la sección eficaz.

La incertidumbre en la eficiencia del algoritmo de identificación de *b*-jets es de un 5-6% [120].

Incertidumbre en la medida de la luminosidad

La medida de la luminosidad en el detector ATLAS tiene una incertidumbre del 3.7% [83–85]. Este efecto se traduce en una incertidumbre del 3.5% en la sección eficaz.

Incertidumbre de los ajustes en la identificación de taus

La incertidumbre debida al propio método se deriva mediante ajustes de una muestra de datos $Z \rightarrow \tau^+\tau^-$ escogida con la misma selección de objetos que en el anásis, pero distintos requerimientos cinemáticos para aumentar la contribución de eventos Z + jets y eliminar los eventos W + jets. La sección eficaz de producción $Z \rightarrow \tau^+\tau^-$ se conoce con gran precisión y, por lo tanto, el error de esta medida nos proporciona la incertidumbre del método de ajustes para la identificación de leptones τ. Los detalles de este cálculo pueden consultarse en el Apéndice C.

Con una incertidumbre del 5% para τ_1 y un 9-10% para τ_3 (ver Tabla 7.11), este error sistemático, junto con el asociado a la identificación de b-jets, tiene la mayor contribución a la incertidumbre total en la sección eficaz.

7.3.5.1 Efecto de las incertidumbres sistemáticas sobre la sección eficaz

El efecto de estas incertidumbres sistemáticas sobre la sección eficaz se evalúa de la siguiente manera. Nuevas muestras MC $t\bar{t}$ se generan variando cada fuente de error en $\pm\sigma$ y de cada una de ellas obtenemos la distribución BDT_j en la región de señal ≥ 1 b-tag. Estas distribuciones tendrán una aceptancia y una forma diferente a la muestra nominal del análisis.

Para los sistemáticos de jets y leptones, obtenidos con el mismo generador que la muestra nominal, evaluamos tanto el cambio en al aceptancia como el efecto del cambio en la forma de la distribución BDT_j, realizando los ajustes con cada muestra de señal y propagando el error en la ecuación 7.20:

$$\varepsilon = \frac{\sigma_{t\bar{t}(SM)}}{S_{MC}} \sqrt{(\Delta S_{meas})^2 + \frac{S^2_{meas}}{S^2_{MC}}(\Delta S_{MC})^2}, \tag{7.16}$$

donde ΔS_{meas} es la variación en el número de eventos de señal medidos respecto a la muestra nominal y ΔS_{MC} es la variación en el número esperado de eventos de señal. Lo que obtenemos es que el cambio en la forma de las distribuciones BDT_j es mínimo en cada muestra y, por lo tanto, la desviación ΔS_{meas} es despreciable y la variación de la aceptancia (ΔS_{MC}) es el efecto dominante en el sistemático de la sección eficaz.

Para los sistemáticos PDF, generador de MC generator y ISR/FSR, propagamos únicamente la incertidumbre en la aceptancia:

$$\varepsilon = \frac{\sigma_{t\bar{t}(SM)} \cdot S_{meas}}{S^2_{MC}} \cdot \Delta S_{MC} \tag{7.17}$$

y no rehacemos los ajustes, ya que cambiar el generador de MC modifica de forma apreciable la forma de las distribuciones BDT_j y este efecto se propaga a las correcciones de MC que aplicamos a los fondos, modificando las distribuciones BDT_j de los fondos en los ajustes y, por lo tanto, el modelo de fondo.

El efecto de cada sistemático en la incertidumbre de la sección eficaz se muestra en % en la Tabla 7.11. Los resultados se muestran por separado para τ_1 y τ_3. La incertidumbre total es del orden del 10-11% para τ_1 y 12% para τ_3. Los sistemáticos dominantes proceden de las incertidumbres en la identificación de taus y b-jets, seguidos de las incertidumbres en la escala de energías de los jets y la introducida por el generador de MC.

Table 7.11: *Errores sistemáticos en la medida de la sección eficaz en el canal $\ell + \tau$ ($\Delta\sigma/\sigma$) usando el método de ajustes en %.*

	$\Delta\sigma/\sigma$ Fitting Method (%)			
	τ_1		τ_3	
	Single p_{T} bins fit	Combined p_{T} fit	Single p_{T} bins fit	Combined p_{T} fit
μ p_T smearing	0.0	0.0	0.0	0.0
μ Trigger SF	−2.7 / +0.4	−2.7 / +0.4	−2.9 / +0.4	−2.9 / +0.4
μ ID SF	0.0	0.0	0.0	0.0
e p_T smearing	±0.2	±0.2	−0.2 / +0.3	−0.2 / +0.5
e energy scale	±0.3	±0.3	−0.2 / +0.3	−0.2 / +0.4
e Trigger SF	±0.2	±0.2	±0.2	±0.2
e ID SF	±1.2	±1.2	±1.2	±1.2
Jet E scale	−2.8 / +2.3	−2.8 / +2.3	−2.7 / +2.6	−2.9 / +2.6
Jet E res.	±0.1	±0.2	±0.4	±0.3
Jet ID eff.	0.0	0.0	0.0	0.0
$b-$tag	−5.0 / +5.7	−5.0 / +5.7	−5.2 / +5.8	−5.1 / +5.8
ISR/FSR	±4.9	±4.2	±4.3	±4.5
Generator	±2.9	±2.9	±1.04	±1.04
Parton Shower	±4.1	±4.1	±0.6	±0.6
τ ID	±5.0	±4.7	±8.8	±9.8
Total	±11.0	±10.3	±12.1	±12.7

7.3.6 Medida de la sección eficaz en el canal $t\bar{t} \to \ell + \tau$

A partir del número de eventos de señal medidos con la técnica de ajustes, podemos obtener la sección eficaz usando la definición estándar:

$$\sigma_{t\bar{t}} = \frac{S_{meas}}{A \cdot \epsilon} \cdot \frac{1}{\mathcal{L}}, \tag{7.18}$$

dónde S_{meas} es la señal medida, A es la aceptancia de la señal, ϵ es la eficiencia del trigger y la reconstrucción de objetos y $\mathcal{L}$ es la luminosidad integrada. La fracción de desintegración $t\bar{t} \to \ell + \tau$ está incluida en A. En la práctica, el modelo de señal de MC provee por construcción la mayoría de los términos de la ecuación anterior:

$$S_{MC} = A\epsilon\mathcal{L}\sigma_{t\bar{t}(SM)}, \tag{7.19}$$

dónde S_{MC} es el número de eventos de señal esperados de MC y $\sigma_{t\bar{t}(SM)}$ es la sección eficaz $t\bar{t}$ en el Modelo Estándar con la que se ha normalizado el MC, es decir, $164.57^{+11.45}_{-15.78}$ pb [28], [29]. Por lo tanto, la fórmula de la sección eficaz queda reducida a:

$$\sigma_{t\bar{t}} = \frac{S_{meas}}{S_{MC}} \cdot \sigma_{t\bar{t}(SM)}. \tag{7.20}$$

En la sección 7.3.4 obtuvimos el número de eventos de señal en la muestra de datos tras la aplicación de la selección de eventos y b-tagging (S_{Fit}) usando el método de ajustes. S_{Fit} incluye

la señal $t\bar{t} \to \ell + \tau$ (S_{meas}), contribuciones de procesos distintos a $t\bar{t}$ con taus reales (principalmente $Z \to \tau\tau$ y una pequeña contribución de single top y dibosones) y una contribución no despreciable de electrones falsamente reconstruidos como leptones τ procedentes de eventos $t\bar{t} \to \ell + e$ and $Z \to ee$. Estas contribuciones se muestran en la Tabla 7.12.

	τ_1		τ_3	
MC Signal	740 ± 6		234 ± 3	
Background τ and e	108 ± 6		28 ± 3	
	Individual fits	Combined fit	Individual fits	Combined fit
Measured τ and e	807 ± 66	802 ± 51	280 ± 52	232 ± 33
Measured τ	699 ± 66	694 ± 51	252 ± 52	204 ± 33
$\frac{S_{meas}}{S_{MC}}$	0.94	0.94	1.07	0.88

Table 7.12: *Taus y electrones medidos a partir de los ajustes. "MC Signal" es el número de leptones τ esperado procedentes de $t\bar{t} \to \ell + \tau$ (S_{MC}). "Background τ and e" representa la contribución de leptones τ de procesos de fondo y electrones, mientras que "Measured τ" es el número de leptones τ medido en el canal $t\bar{t} \to \ell + \tau$ (S_{meas}). La fracción $\frac{S_{meas}}{S_{MC}}$ se calcula tras la substracción de los fondos.*

El número de eventos de señal medidos se obtiene finalmente como:

$$S_{meas} = S_{Fit} - N^{bkg}_{e,\tau}. \tag{7.21}$$

Los resultados finales para la sección eficaz se dan para τ_1 y τ_3 por separado. Los errores sistmáticos se corresponden con la Tabla 7.11. Los resultados finales para la medida de la sección eficaz en el canal $t\bar{t} \to \ell + \tau$ usando la técnica de ajustes son:

$$\sigma(\tau_1) = 155 \pm 15 \text{ (stat.)} \pm 17 \text{ (syst.)} \pm 5 \text{ (lumi.) pb.} \tag{7.22}$$

$$\sigma(\tau_3) = 176 \pm 62 \text{ (stat.)} \pm 21 \text{ (syst.)} \pm 6 \text{ (lumi.) pb.} \tag{7.23}$$

y con la técnica de ajustes combinados:

$$\sigma(\tau_1, Comb.) = 154 \pm 12 \text{ (stat.)} \pm 16 \text{ (syst.)} \pm 5 \text{ (lumi.) pb.} \tag{7.24}$$

$$\sigma(\tau_3, Comb.) = 143 \pm 23 \text{ (stat.)} \pm 18 \text{ (syst.)} \pm 5 \text{ (lumi.) pb.} \tag{7.25}$$

7.4 Conclusiones

El trabajo desarrollado en esta tesis se ha centrado en la medida de la sección eficaz de producción $t\bar{t}$ en el canal dileptónico con leptón tau ($t\bar{t} \to \ell + \tau$) utilizando 2.05 fb^{-1} de datos de colisiones tomados por el detector ATLAS durante 2011.

Los leptones tau juegan un papel importante en búsquedas del bosón de Higgs. Un bosón de Higgs neutro estándar puede producirse asociado a un par top-antitop, $t\bar{t}H^0$, y desintegrarse a dos leptones tau: $H^0 \to \tau\tau$ aproximadamente un 10% de los casos en el rango de baja masa del bosón de Higgs. Medir este canal de desintegración proporcionaría evidencia directa del

acoplamiento de fermiones al bosón de Higgs. Bosones de Higgs cargados con masa menor a la del quark top pueden producirse a través de desintegraciones de quarks top: $t \to H^+ b$. Bosones de Higgs cargados masivos también pueden producirse en asociación con un quark top: tH^+. En algunos escenarios, el Higgs cargado se desintegra predominantemente a un tau y un neutrino, produciendo un exceso en el canal $\ell + \tau$ sobre otros canales dileptónicos. Por lo tanto, cualquier desviación en la medida de la sección eficaz $t\bar{t} \to \ell + \tau$ respecto al valor previsto por el Modelo Estándar constituiría un indicio experimental de la existencia de un bosón de Higgs cargado.

La selección de eventos $t\bar{t} \to \ell + \tau$ se realiza de acuerdo a la topología de su estado final en el detector: b-jets de alta E_{T}, un leptón de alto p_{T}, un τ que decae hadrónicamente y elevada $E_{\mathrm{T}}^{\mathrm{miss}}$. El fondo dominante tras la selección prodece the eventos $t\bar{t} \to \ell$ + jet donde un jet ha sido misidentificado como un tau. La única forma de diferenciar señal y fondo es mediante la identificación de leptones tau. Para ello utilizamos una variable discriminante, BDT_j, obtenida mediante técnicas multivariadas de árboles de decisión entrenada especialmente para separar taus que decaen hadrónicamente de otro tipo de jets.

Las contribuciones de señal y fondo en la muestra de datos se estiman mediante un ajuste χ^2 de la distribución BDT_j de los taus en la muestra de datos con un patrón para la señal obtenido a partir de MC y patrones para los fondos obtenidos a partir de datos, para minimizar la dependencia con la simulación. Utilizamos dos patrones para modelizar los fondos procedentes de quarks ligeros y gluones, que son la fuente dominante de falsos taus en la muestra $\ell + \tau$ tras la selección. Los parámetros del ajuste son las cantidades de señal y fondo, mientras que la forma de las distribuciones es fija.

La sección eficaz $t\bar{t} \to \ell + \tau$ obtenida para τ_1 es:

$$\sigma(\tau_1) = 154 \pm 12 \text{ (stat.)} \pm 16 \text{ (syst.)} \pm 5 \text{ (lumi.) pb.} \tag{7.26}$$

y para τ_3:

$$\sigma(\tau_3) = 143 \pm 23 \text{ (stat.)} \pm 18 \text{ (syst.)} \pm 5 \text{ (lumi.) pb.} \tag{7.27}$$

Esta medida mejora la precisión de los resultados obtenidos en los experimentos DØ, CDF y CMS, con incertidumbres del 30% [63], 25% [65] y 18% [66], respectivamente, y la última medida publicada de ATLAS [67].

Medidas previas de ATLAS en el canal $\ell + \tau$ se obtuvieron mediante un ajuste χ^2 de distribuciones BDT_j OS-SS y se verificaron utilizando una técnica matricial descrita en el Apéndice A. La motivación para utilizar la substracción OS-SS era eliminar el fondo de gluones y mantener los quark ligeros como el único fondo, mientras que la señal permanecía intacta. El fondo de quarks ligeros se estimaba a partir de datos con una selección cinemática similar a la de la señal, pero sin presencia de b-jets. La contribución de taus reales se eliminaba del patrón de fondo substrayendo la predicción de MC.

Usando las muestras OS en lugar de OS-SS en nuestro análisis aumentamos la estadística de los patrones utilizados en los ajustes y reducimos el error estadístico en al medida de la sección eficaz. También mejoramos la descripción de los fondos contruyendo un modelo que refleja apropiadamente la composición de los jets misidentificados como taus. Los errores sistemáticos limitan la preción de la medida y se han mantenido en los niveles de análisis anteriores.

El valor observado de la sección eficaz concuerda con el valor predicho por el Modelo Estándar de $164.57^{+11.45}_{-15.78}$ pb y es consistente con las medidas realizadas en el resto de canales de desintegración por ATLAS [40] [41] [42] [43] y CMS [45] [46] [47], como se puede observar en la

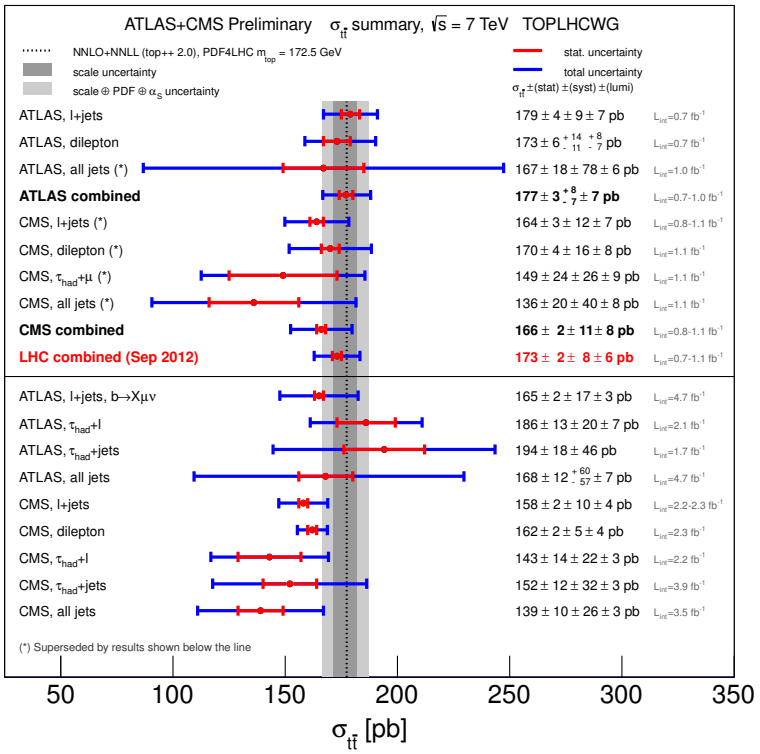


Figure 7.8: *Medida de la sección eficaz de producción $t\bar{t}$ en varios canales de desintegración con los experimentos ATLAS y CMS.*

Figura 6.1. La ausencia de discrepancias con el Modelo Estándar, con las preciones actuales, mantiene activa la búsqueda de nueva física.

La técnica desarrollada en esta tesis posee un considerable potencial para alcanzar grandes precisiones al aplicarse a muestras de alta estadística, como la muestra completa de datos tomados por el LHC durante los años 2011 y 2012 y los datos de futuras colisiones con una energía de 13-14 TeV que se producirán durante el Run II.

Esta técnica se ha aplicado a los 5.08 fb^{-1} de datos tomados por ATLAS en 2011 para la medida de la sección eficaz de producción $t\bar{t} \to \ell + \tau$ [153], medida de las las fracciones de desintegración de pares de quark top en leptones y jets [154], orientada a búsquedas de física más allá del Modelo Estándar, y en la medida de la incertidumbre de identificación de leptones τ, que es una de las mayores fuentes de error sistemático en el análisis. Con más del doble de estadística en la muestra de datos éste sistemático se ha reducido más de un 40%, pasando de un 5%(9%) a un 2.4%(5.1%) para $\tau_1(\tau_3)$ [155].

La técnica de ajustes podrá además perfeccionarse con bines de p_T más finos, de precisión igual o mayor, de modo que la evolución de los cambios en la forma de la variable BDT_j con el p_T del τ podrán ser monitorizados con más detalle, mejorando la descripción de los fondos.

Appendices

A

$t\bar{t} \to \ell + \tau$ cross section measurement cross check: the Matrix Method

A cross check of the cross section measurement has been done using the so-called matrix method developed in references [137] and [67]. Relying on the OS-SS technique described in Section A.1 below to simplify the background model, the matrix method solves a simple system of equations based on the tau identification efficiency for signal and background at a cut value of $BDT_j > 0.7$. The signal efficiency is estimated from MC and the fake rate is estimated using a W+ jets selection from data, enriched in light-quark jets.

A.1 OS-SS analysis technique

As we have seen in chapter 5, after applying all the kinematic selections and the b-tagging requirement, the main contributing background is the top quark pair production in the lepton plus jets channel in which the τ candidate come from jets misidentified as hadronic τ decays. Therefore, τ lepton identification is the critical discriminator between signal and background.

The jet origin can strongly influence the τ lepton fake rate probability. Due to their narrow shower width and lower track multiplicity, light-quark jets have a higher probability of faking a τ lepton than other jet types. Therefore, the BDT_j distributions have a strong dependence on the jet type, which means that the tau identification does not equally perform against all types of jet. According to the MC simulation, shown in Table A.1, 16% of the fake τ candidates come from gluon jets, 78% from light-quark jets and 6% from b-jets in the case of τ_1 candidates, while in the case of τ_3 candidates 17% of the fake τ candidates come from gluon jets, 67% from light-quark jets, and 8% from b-jets.

Therefore it will be crucial to build a background model which properly reflects the jet composition in order to correctly estimate the fake τ contamination in the signal region. Deriving

this background model from control regions in data rather than MC simulation is preferable in order to avoid systematic effects related to jet composition in the MC models, but it is not an easy task since we do not know the right jet composition in data.

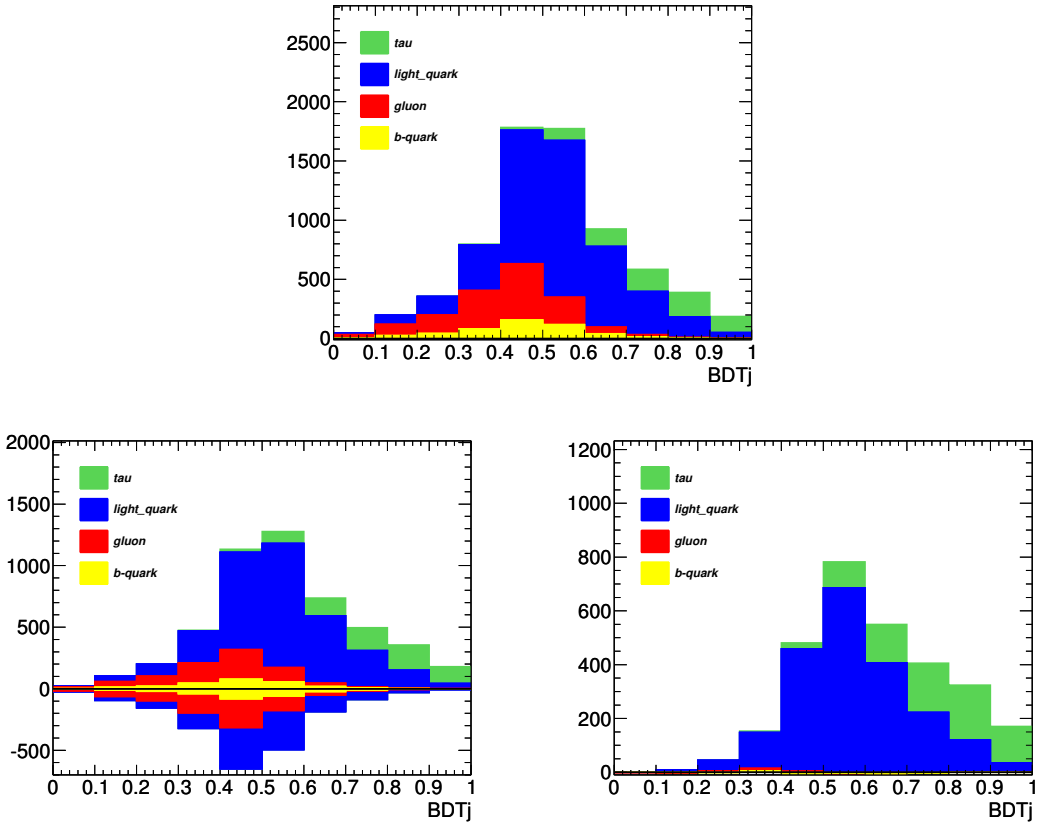


Figure A.1: *BDT_j distribution for b-tag region by truth type of the reconstructed τ_1 object. The top plot shows all OS+SS events. Left Bottom plot shows OS in the positive y axis and SS in the negative y axis. Real τ candidates contribute only to the OS events, while fake τ candidates contribute to both OS and SS events. The right bottom plot shows the OS-SS events, which are mainly real taus and light-quark jets.*

The BDT_j distribution built with all MC tau candidates, signal and background, in the ≥ 1 b-tag signal region described in section 5.2 is shown in Figure A.1. The tau candidates are categorized by the truth particle matched to the tau: real taus, b-quarks, gluon and light-quark jets. The contribution from electrons or muons faking the τ is negligible (as can be seen in Table A.1) and, therefore, is not included. The bottom left plot shows the same BDT_j distributions split into two separate distributions, depending on the charge correlation between the lepton and the τ candidate: positive y axis for opposite sign charged leptons (OS) and negative y axis for same sign charged leptons (SS). On the bottom right plot the OS-SS BDT_j distributions are

displayed.

From these plots it can be observed that the BDT_j distributions for gluon jets are charge symmetric, since they have the same shape and the same number of τ candidates matched to gluons in the SS and OS samples. Thus, the contribution of fake τ candidates from gluon jets can be removed by subtracting the BDT_j distribution for SS events from the corresponding distribution for OS events. The QCD multi-jet background also cancels, as can be seen in Table 5.5.

Each sample is expected to have an equal contribution from b-jets and $\bar{b}$-jets. Therefore, the small b-jet component in the data sample is also removed by the OS-SS technique, within statistical fluctuations, as can be seen in Figure A.1.

The only jet types remaining in the data samples, besides true τ leptons, are light-quark jets. Therefore, using the OS-SS subtraction, we can perform measurements in the signal region with the assumption that all fake τ leptons come from light-quark jets. This technique reduces systematic errors due to b quarks and gluon contaminating the signal.

The OS-SS subtraction technique has been validated using the MC simulation as can be seen if Figure A.1. The bottom right plot shows OS-SS distribution, where the τ candidates coming from gluon jets and b-jets are almost cancelled out and the remnant contributions are light-quark jets and the true τ candidates from signal events. QCD multi-jet background events, although not included in the plot, also cancel out with the OS-SS subtraction (see Table 5.5) since the fake τ candidates originating from the QCD multi-jet background are not related to the lepton charge and the BDT_j distribution built with them will be charge symmetric.

A.2 Data-derived light quark background

To model the background in the ≥ 1 b-tag signal region, a source of light-quark jets is needed. We can obtain a sample of W+1 jet events from data, which is rich in quark jets, doing the following selection:

- $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV , to reduce QCD multi-jet background.
- $40 < M_{\mathrm{T}}(E_{\mathrm{T}}^{\mathrm{miss}}, \ell) < 100$ GeV, to reject Z+jets events.
- One reconstructed lepton (electron or muon).
- Only one jet in the event, overlapping with the τ candidate. This vetoes b-jets from the event since τ candidates are preselected applying overlap removal with b-tagged jets.

Applying this selection and requiring the lepton and τ candidate to be oppositely charged (OS), we obtain a sample where 90% of the fake taus come from quark jets and the 10% remaining from gluon jets. This sample is labeled W+1 jet (OS). By applying identical selections, but reversing the charge sign requirement, we obtain a second data sample which we label W+1 jet (SS). The selection ensures that the fakes from this sample are still quark dominated, with about 2/3 of the fake τ leptons coming from quark jets and 1/3 from gluon jets.

Since the gluon component of the τ fakes is charge symmetric, it should have the same shape in the W+1 jet (SS) sample as in the W+1 jet (OS) sample. The influence of gluon fakes can then be removed by performing the subtraction OS-SS. The expected jet compositions before and after OS-SS subtraction are shown in Table A.1.

Table A.1: *Composition of all τ candidates in MC events, after requiring the W + 1 jet and $\geq$ 1 b-tag selections detailed in the text. OS (SS) stands for the opposite (same) charge sign between e or μ and τ canndidate (which is a jet selected as τ in this study).*

Jet Composition of Jet-to-τ Fake Candidates

τ_1	gluon	quark	b	τ	e	μ
W + 1 jet (SS)	30%	70%	0%	0%	0%	0%
W + 1 jet (OS)	10%	90%	0%	0%	0%	0%
W + 1 jet (OS-SS)	1%	99%	0%	0%	0%	0%
$\geq$ 1 b-tag (SS)	33%	55%	12%	0%	0%	0%
$\geq$ 1 b-tag (OS)	13%	65%	5%	16%	1%	0%
$\geq$ 1 b-tag (OS-SS)	1%	72%	0%	27%	1%	0%
τ_3	gluon	quark	b	τ	e	μ
W + 1 jet (SS)	35%	65%	0%	0%	0%	0%
W + 1 jet (OS)	17%	83%	0%	0%	0%	0%
W + 1 jet (OS-SS)	2%	98%	0%	0%	0%	0%
$\geq$ 1 b-tag (SS)	40%	47%	13%	0%	0%	0%
$\geq$ 1 b-tag (OS)	24%	66%	8%	2%	0%	0%
$\geq$ 1 b-tag (OS-SS)	0%	95%	0%	5%	0%	0%

The W+1 jet (OS-SS) control region is, thus, dominated by light-quark jets and it is an ideal candidate to calculate the τ fake rate in the $\geq$ 1 b-tag data sample. However, fake τ candidates coming from W+1 jet events are kinematically different from those coming from $t\bar{t}$ events. This affects the tau identification variables which results in the OS-SS BDT_j distributions in both regions not being identical and, therefore, the fake rates are different.

Two different approaches can be used to construct the light-quark background to estimate the τ fake rate. One is to apply corrections to the shape of the W+1 jet BDT_j based on MC. The other approach, which we apply in this cross-check analysis, aims to avoid applying MC based corrections to the data derived background. This is done splitting the background into bins of some variable within which the shapes of the BDT_j distributions of the background model are close to those from the $\geq$ 1 b-tag MC background, constructed with all non signal MC processes after applying the selection that define the $\geq$ 1 b-tag signal region. This variable is the electromagnetic fraction (f_{em}) of the τ candidate. Figures A.2 and A.3 show the BDT_j distributions for the W+1 jet sample (in black) and the $\geq$ 1 b-tag MC background (in red) for both τ_1 and τ_3 candidates.

This approach avoids the use of MC corrections in data, but assumes the data and MC simulation behave similarly as function of the binning variable.

A.3 Signal and bakground estimation using the Matrix Method

The matrix method divides τ distributions into two regions based on a BDT_j cut. All τ candidates are considered loose, and all τ leptons with $BDT_j > 0.70$ are considered tight. In a given region the number of events in the loose sample ($N_{\text{data}}^{\text{loose}}$) is given by

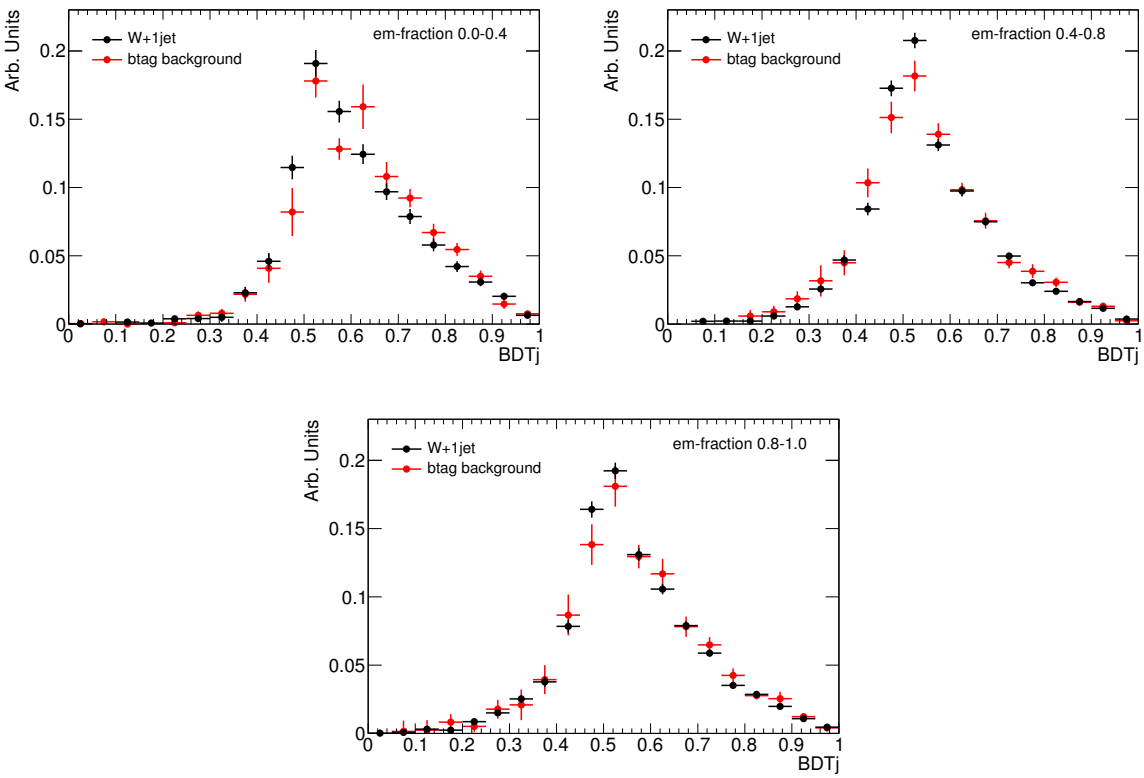


Figure A.2: *OS-SS BDT_j from MC for fake τ_1s for W+1 jet control region (black) and $\geq$ 1 b-tag background (red) in bins of f_{em}. Once binned by f_{em}, the W+1 jet distribution agrees well with the signal region background.*

$$N_{\text{data}}^{\text{loose}} = N_{\text{fake}}^{\text{loose}} + N_{\text{real}}^{\text{loose}}, \tag{A.1}$$

where the real subscript denotes events with a real τ and the fake superscript denotes events with a fake τ. The probability that the loose selection passes the tight cut, for both real and fake τ leptons, is defined as:

$$\epsilon_{\text{real}} = \frac{N_{\text{real}}^{\text{tight}}}{N_{\text{real}}^{\text{loose}}}, \quad \epsilon_{\text{fake}} = \frac{N_{\text{fake}}^{\text{tight}}}{N_{\text{fake}}^{\text{loose}}}. \tag{A.2}$$

ϵ_{real} is derived from MC (including all processes that contribute a τ lepton or an electron misidentified as a τ lepton) and ϵ_{fake} is measured using the W+1 jet OS-SS BDT_j distribution from data. Values of ϵ_{real} and ϵ_{fake} are measured separately for the three f_{em} bins (Fig. A.4).

The number of observed tight events is given by

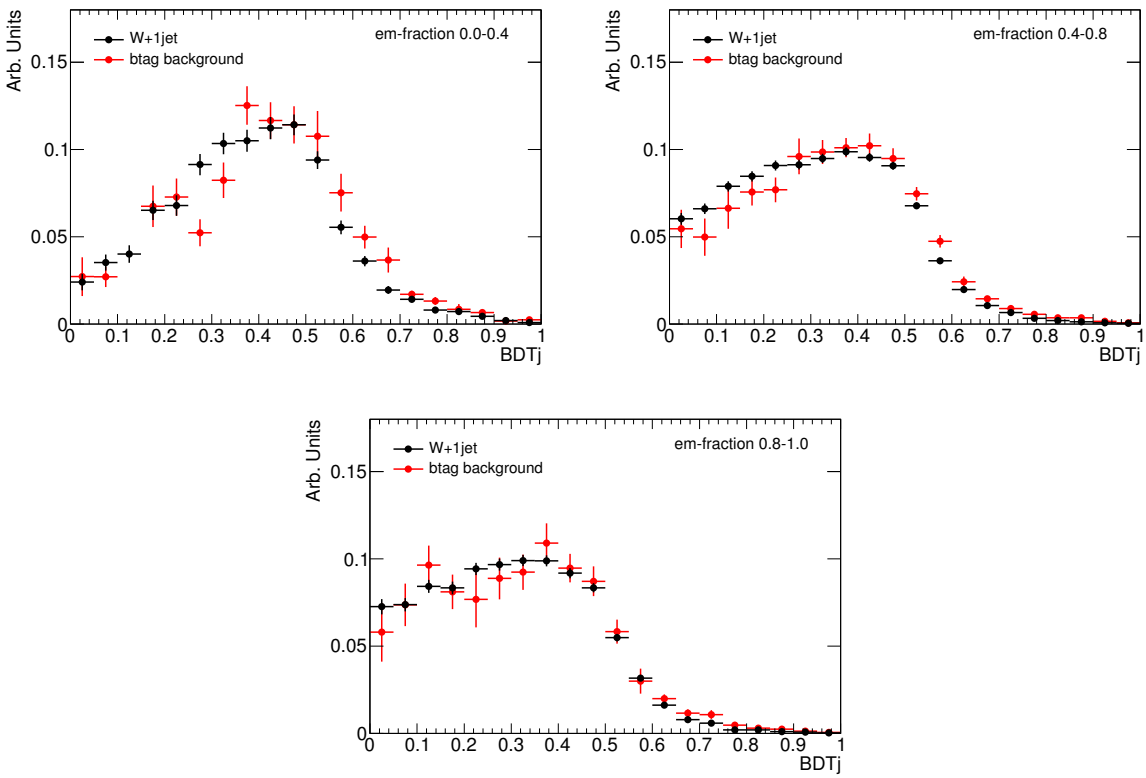


Figure A.3: *OS-SS BDT$_j$ from MC for fake τ_3s for W+1 jet control region (black) and $\geq$ 1 b-tag background (red) in bins of f_{em}. Once binned by f_{em}, the W+1 jet distribution agrees well with the signal region background.*

$$N_{\text{data}}^{\text{tight}} = N_{\text{fake}}^{\text{tight}} + N_{\text{real}}^{\text{tight}}. \tag{A.3}$$

Solving the system of equations A.1, A.1 and A.1, we obtained the background prediction:

$$N_{\text{fake}}^{\text{tight}} = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}}(N_{\text{data}}^{\text{loose}} * \epsilon_{\text{real}} - N_{\text{data}}^{\text{tight}}), \tag{A.4}$$

and the signal prediction:

$$N_{t\bar{t}\,\tau's}^{\text{tight}} = N_{\text{data}}^{\text{tight}} - N_{\text{fake}}^{\text{tight}} - N_{\text{lepton}}^{\text{tight}} - N_{\text{othertau}}^{\text{tight}}, \tag{A.5}$$

where $N_{\text{othertau}}^{\text{tight}}$ is the contribution from non $t\bar{t}$ process that produce true τ leptons and $N_{\text{lepton}}^{\text{tight}}$ that of electrons faking a τ candidate. All distributions used are after the OS-SS subtractions.

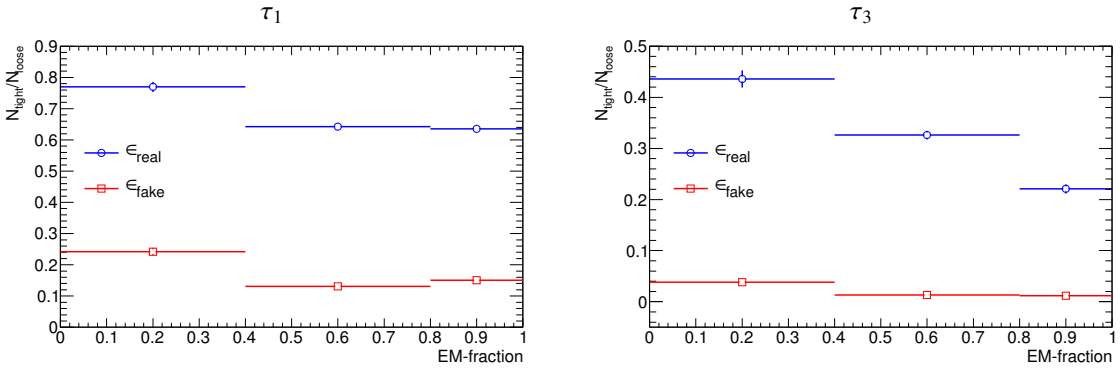


Figure A.4: *ϵ_{real} from MC for the ≥ 1 b-tag signal region (blue) and ϵ_{fake} measured in the W+1 jet data control region (red). Left is τ_1 and right is τ_3.*

A.3.1 Testing ϵ_{real} with $Z \rightarrow \tau\tau$ samples

The matrix method requires two terms to be modelled accurately: ϵ_{real} and ϵ_{fake}. Testing ϵ_{real} is of particular importance, since it is taken solely from MC. This section describes the use of a $Z \rightarrow \tau\tau + 0$ jet selection from data to assess the reliability of ϵ_{real}.

$Z \rightarrow \tau\tau$ events, where one τ decays to a lepton and the other decays hadronically, are selected from data with the following requirements:

- exactly one lepton (e or μ) and one τ candidate
- only one jet in the event, overlapping with the τ candidate
- the transverse mass of the lepton and $E_{\text{T}}^{\text{miss}}$ is required to be $M_T(E_{\text{T}}^{\text{miss}}, \ell) < 20$ GeV, to reject W + jets processes

In order to apply the matrix method, $N_{\text{data}}^{\text{loose}}$ and $N_{\text{data}}^{\text{tight}}$ values were measured in the $Z \rightarrow \tau\tau + 0$ jet data as a function of f_{em} after performing the OS-SS subtraction. The ϵ_{real} term was derived using the MC expectation for real τ leptons surviving the $Z \rightarrow \tau\tau + 0$ jet selection, while ϵ_{fake} was taken from the $W + 1$ jet background model in data. Both terms were measured as a function f_{em}. The results of the matrix method applied in this region are shown, after integrating over f_{em}, in Table A.2.

The τ_1 results ($N_{\text{real}}^{\text{tight}}$) are within 3% of MC expectation, which is well within the expected τ uncertainty of 5.0%. The τ_3 events agree within 10% which is statistically compatible with τ_3 uncertainty of 7.1% [104]. The effect on the background prediction in the signal region caused by varying ϵ_{real} by $\pm$ 3% and $\pm$ 10% is small and shown in Table A.3. Since this effect is completely covered by the τ uncertainty, no additional systematic is applied.

A.3.2 Results of the matrix method

Since the matrix method is solved independently in bins of f_{em} it is possible to check whether the observed shapes agree with MC expectation. The results as a function of f_{em} are displayed

τ_1	$e + \tau$	$\mu + \tau$	Comb.
Events ($N_{\text{data}}^{\text{tight}}$)	3879	5238	9117
Measured Background ($N_{\text{fake}}^{\text{tight}}$)	337 ± 97	580 ± 78	918 ± 117
Measured Signal ($N_{\text{real}}^{\text{tight}}$)	3542 ± 132	4658 ± 125	8199 ± 176
MC expected	3718	4309	8027
τ_3	$e + \tau$	$\mu + \tau$	Comb.
Events ($N_{\text{data}}^{\text{tight}}$)	444	948	1392
Measured Background ($N_{\text{fake}}^{\text{tight}}$)	86 ± 10	164 ± 12	248 ± 15
Measured Signal ($N_{\text{real}}^{\text{tight}}$)	358 ± 38	784 ± 46	1144 ± 59
MC expected	451	819	1269

Table A.2: *Matrix Method results in the $Z \to \tau\tau$ + 0 jet control region. Events are integrated over bins of f_{em}.*

Table A.3: *Background measured with data in the $\geq$ 1 b signal region with the matrix method after varying ϵ_{real} within uncertainty. The uncertainty was measured using the $Z \to \tau\tau$+ 0 jet selection, as shown in Table A.2.*

τ_1	-3%	0	+3%
Jet Background	229 ± 21	234 ± 22	239 ± 22
τ_3	-10%	0	+10%
Jet Background	54 ± 4	54 ± 4	54 ± 4

with the SM prediction in Figure A.5 and Figure A.6. All curves agree within errors.

Table A.4 shows the total number of signal events obtained with the matrix method (adding up the results on each f_{em} bin) for τ_1 candidates, as well as the measured light-quark jet background and the expected values for signal and e/τ background. The numbers are in good agreement with MC expectations. Same results are obtained for the τ_3 case (Table A.5). The deviations from the numbers obtained with the fitting method (section 5.5.2) are due to the fact that only one τ candidate per event (the leading τ) is considered in the matrix method, whereas all tau candidates are used to fill the templates in the fitting method.

A.4 $t\bar{t}$ cross section measurement with the Matrix Method

The source of the systematic uncertainties arises from MC expectation on acceptance of tight τ candidates which pass $BDT_j > 0.7$. The impact of systematic uncertainties is estimated by shifting a given parameter within its uncertainty and redoing the full analysis chain. Individual sources of systematics uncertainties are detailed in section 5.6 and their impact in their final cross section measurement is shown in Table A.6.

The cross section is derived from the number of measured signal events with the matrix method using the standard definition as described in section 5.7:

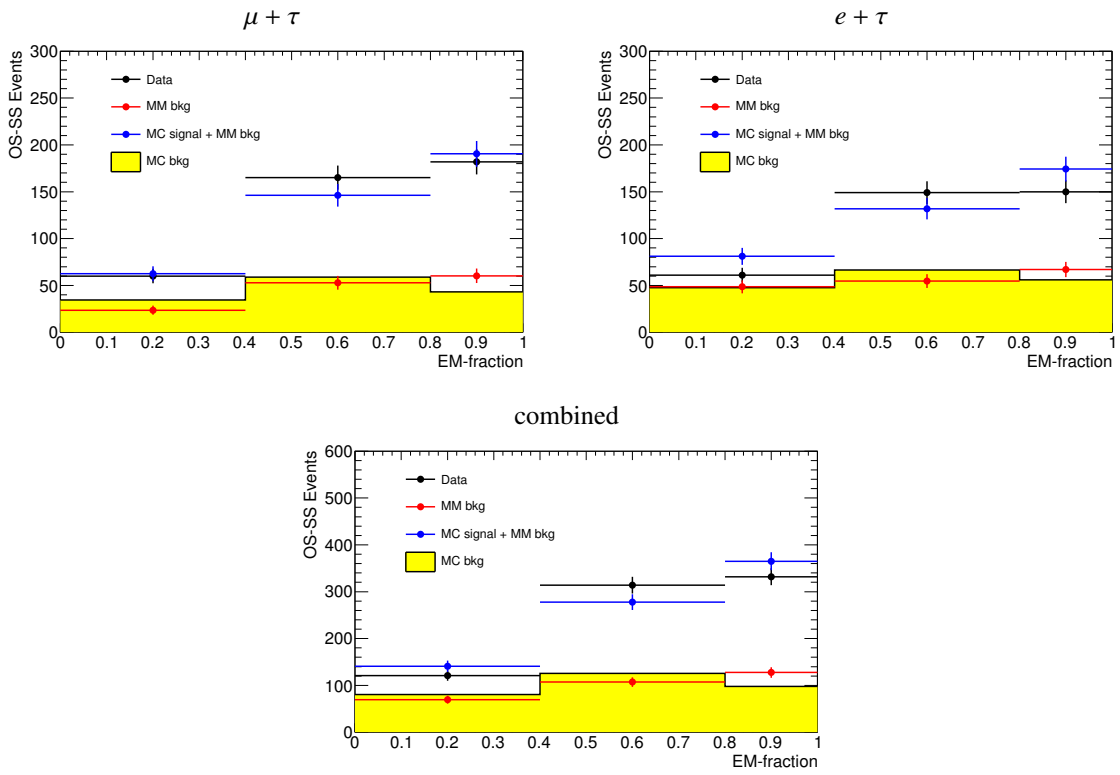


Figure A.5: *τ_1 results of the matrix method in the signal region (e, μ). Yellow MC prediction is normalized to the matrix method prediction (red).*

Table A.4: *Event yields in the signal region for τ_1. The "Expected τ background" and "Expected e background" are taken from MC and are subtracted from the "Total Signal" to obtain the final "Measured $t\bar{t} \to \ell + \tau$ signal" result. Events are integrated over bins of f_{em}.*

	$\mu + \tau$	$e + \tau$	Comb.
Events ($N_{\text{data}}^{\text{tight}}$)	407	360	767
Measured jet background ($N_{\text{fake}}^{\text{tight}}$)	101 ± 15	133 ± 16	234 ± 21
Total Signal ($N_{\text{real}}^{\text{tight}}$)	306 ± 28	227 ± 27	533 ± 39
Expected τ background	28	27	56
Expected e background	7	9	17
Measured $t\bar{t} \to \ell + \tau$ signal	270 ± 28	190 ± 27	460 ± 39
Expected $t\bar{t} \to \ell + \tau$ signal	263	217	480

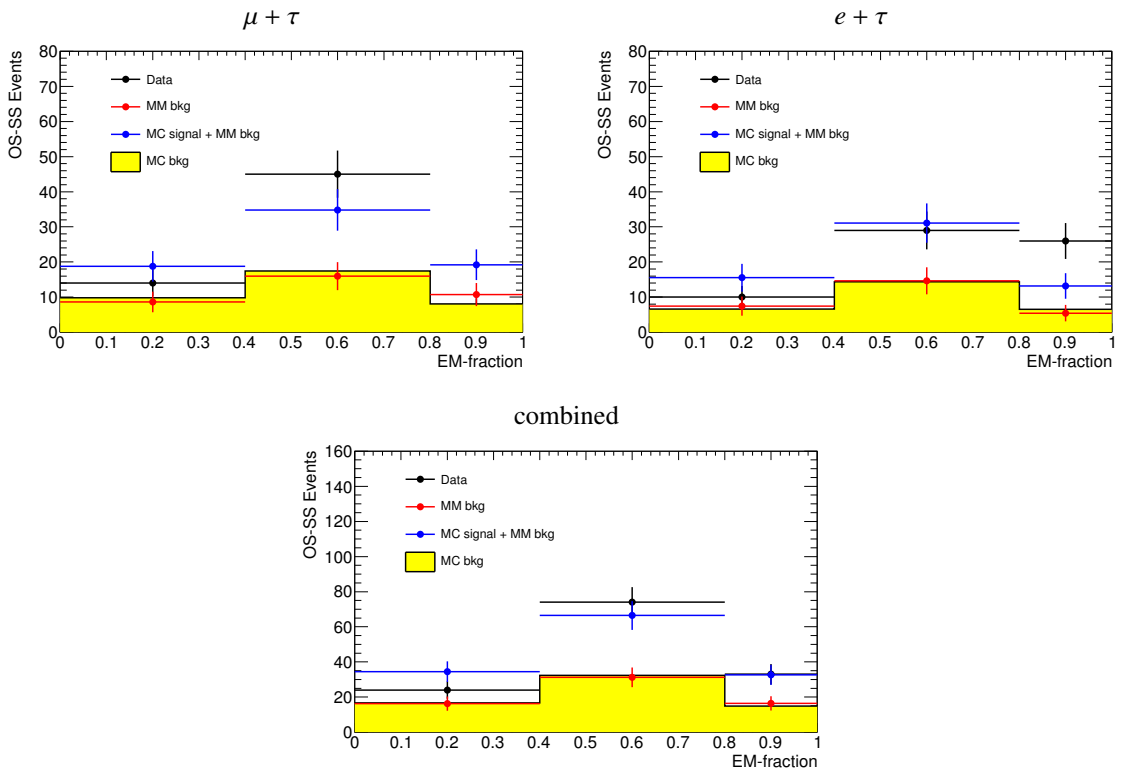


Figure A.6: *τ_3 results of the matrix method in the signal region (e, μ). Yellow MC prediction is normalized to the matrix method prediction (red).*

Table A.5: *Event yields in the signal region for τ_3. The "Expected τ background" and "Expected e background" are taken from MC and are subtracted from the "Total Signal" to obtain the final "Measured $t\bar{t} \to \ell + \tau$ signal" result. Events are integrated over bins of f_{em}.*

	$\mu + \tau$	$e + \tau$	Comb.
Events ($N_{\text{data}}^{\text{tight}}$)	66	65	131
Measured jet background ($N_{\text{fake}}^{\text{tight}}$)	30 ± 3	24 ± 3	55 ± 3
Total Signal ($N_{\text{real}}^{\text{tight}}$)	36 ± 10	41 ± 10	77 ± 14
Expected τ background	5	3	8
Expected e background	0	0	0
Measured $t\bar{t} \to \ell + \tau$ signal	31 ± 10	37 ± 10	68 ± 14
Expected $t\bar{t} \to \ell + \tau$ signal	37	32	70

Table A.6: *Systematic uncertainties on the total cross section ($\Delta\sigma/\sigma$) measurement for the matrix method in % for $\mu + \tau$ and $e + \tau$ channels.*

	$\Delta\sigma/\sigma$ MM	
Source	$\mu + \tau$	$e + \tau$
μ p_T res.	±0.1	±0.0
μ (ID/Trig.)	±2.8	±0.0
e p_T res.	±0.0	−0.7 / +0.8
e (ID/Trig.)	±0.0	±2.5
Jet E scale	−1.8 / +2.3	±3.0
Jet E res.	±0.6	±0.5
Jet ID eff.	±0.0	±0.0
b−tag	−4.5 / +5.8	−5.0 / +5.7
ISR/FSR	±6.4	±6.7
MC Generator	±0.8	±5.1
Parton Shower	±4.5	±3.9
τ_1 ID	±5.0	±5.0
τ_3 ID	±7.1	±7.1

$$\sigma_{t\bar{t}} = \frac{S_{meas}}{S_{MC}} \cdot \sigma_{t\bar{t}(SM)}. \tag{A.6}$$

where S_{meas} is the measured signal, S_{MC} is the number of signal events expected by MC and $\sigma_{t\bar{t}(SM)}$ is the standard model $t\bar{t}$ cross section by which the MC was normalized, i.e., $164.57^{+11.45}_{-15.78}$ pb [28], [29]. These contributions are shown in table A.7 for τ_1 and τ_3 in the two lepton plus tau channels.

Table A.7: *Measured and MC numbers of τ and e objects in the ≥ 1 b-tags region from the matrix method using the W+1 jet background model. The "Measured τ and e" are the total number of τ and e objects found by the matrix method calculations, including those from processes other than $t\bar{t} \to \ell + \tau$, such as $t\bar{t} \to e + \ell$ and $Z \to \tau\tau$. The "MC Signal" are the expected number of τ leptons from $t\bar{t} \to \ell + \tau$, while the "Background τ and e" are the number of objects from these non $t\bar{t} \to \ell + \tau$, estimated from MC. $\frac{S_{meas}}{S_{MC}}$ is the ratio of measured to MC signal after subtracting these background τ and e numbers from the measured signal. All numbers are for OS-SS distributions.*

Channel	Measured τ and e	MC Signal	Background τ and e	$\frac{S_{meas}}{S_{MC}}$
$\mu + \tau_1$	306	263	35	1.02
$\mu + \tau_3$	36	37	5	0.84
$e + \tau_1$	227	217	36	0.88
$e + \tau_3$	41	32	3	1.15

The final cross sections measured using the matrix method are given in Table A.8 separately for τ_1 and τ_3 in the $e + \tau$ and $\mu + \tau$ channels, respectively. The systematic uncertainties are taken from table A.6.

Table A.8: *Measured cross section using the matrix method in the τ_1 and τ_3 channels*

	$\mu + \tau$
τ_1	169 ± 17 (stat.) $\pm^{19}_{18}$ (syst.) ± 6 (lumi.)
τ_3	135 ± 45 (stat.) $\pm^{17}_{16}$ (syst.) ± 5 (lumi.)
	$e + \tau$
τ_1	144 ± 21 (stat.) ± 18 (syst.) ± 5 (lumi.)
τ_3	191 ± 51 (stat.) ± 25 (syst.) ± 7 (lumi.)

B

W + 1 jet composition studies

As has been shown on chapter 5, the shape of the fake τ BDT_j distributions depend strongly on the type of jet. Estimating the background for a particular channel thus requires knowing the precise mixture of jet types that contribute to it. MC studies showed that after applying the event selection in section 5.2 and the b-tagging requirement, customized for the $t\bar{t}$ lepton plus tau channel, the jet composition of fake τ leptons was dominated by light-quark and gluon jets.

A light-quark jet dominated sample is obtained by applying the following W + 1 jet selection to data:

- only one isolated lepton (electron or muon) as described in 3.2
- only one reconstructed τ, after applying overlap removal with b-jets
- $E_{\mathrm{T}}^{\mathrm{miss}} \geq 30$ GeV to reduce the QCD multi-jet background
- $40 \leq M_{\mathrm{T}}(E_{\mathrm{T}}^{\mathrm{miss}}, \ell) \leq 100$ GeV to reject Z + jets events.
- If there is only one jet in the event and it overlaps with the τ candidate the event will be classified as a W + 1 jet event. If there are two jets in the event and the τ candidate overlaps with the low pt one then the event will be considered a W + 2 jets event.

The gluon jets BDT_j distribution is obtained from the W + 1 jet and W + 2 jets samples as described in Section 5.4. W + 1 jet and gluon BDT_j distributions were thus used to estimate the background in the data sample.

If we compare the background templates derived from data with those derived from MC we can see that Monte Carlo simulation does not reproduce well the BDT_j distributions in a data sample. In Figure B.1 we can see that the shape of the OS and SS W + 1 jet templates are not in good agreement but the derived gluon templates match much more closely. There are two possible reasons for the OS and SS mismatch between data and MC:

1. MC fails to describe properly the light-quark jet fragmentation when the track multiplicity is low
2. The mixture of light-quark jets and gluons is significantly different in the data and the MC samples

In order to derive the gluon content as a percentage of the total number of OS events in the W + 1 jet dataset:

$$N_{evt}(\text{Gluons}) = p \cdot N_{evt}(\text{OS1}), \tag{B.1}$$

we make use of the fact that the gluon contribution to OS and SS W + 1 jet events is the same:

$$N_{evt}(\text{OS1}) = N_{evt}(\text{OS}_q) + N_{evt}(\text{Gluons}), \tag{B.2}$$

$$N_{evt}(\text{SS1}) = N_{evt}(\text{SS}_q) + N_{evt}(\text{Gluons}), \tag{B.3}$$

where $N_{evt}(\text{OS}_q)$ and $N_{evt}(\text{SS}_q)$ are the number of events in the W + 1 jet OS and SS samples where the jet originated by the fragmentation of a light quark is faking the tau candidate. We can, therefore, express the number of light-quark jets as:

$$N_{evt}(\text{OS}_q) = N_{evt}(\text{OS1}) - N_{evt}(\text{Gluons}) = (1-p) \cdot N_{evt}(\text{OS1}), \tag{B.4}$$

$$N_{evt}(\text{SS}_q) = N_{evt}(\text{SS1}) - N_{evt}(\text{Gluons}) = N_{evt}(\text{SS1}) - p \cdot N_{evt}(\text{OS1}). \tag{B.5}$$

Assuming that the ratio

$$R = \frac{N_{evt}(\text{OS}_q)}{N_{evt}(\text{SS}_q)} = \frac{(1-p) \cdot N_{evt}(\text{OS1})}{N_{evt}(\text{SS1}) - p \cdot N_{evt}(\text{OS1})} \tag{B.6}$$

is constant in al the range of p_T for data and MC, we can compute it initially for all the range $20 < p_T < 200$ GeV using the hypothesis of $p = 10\%$ for τ_1 and $p = 17\%$ for τ_3 in MC. Then, in each p_T bin for both data and MC we can compute the value of p as

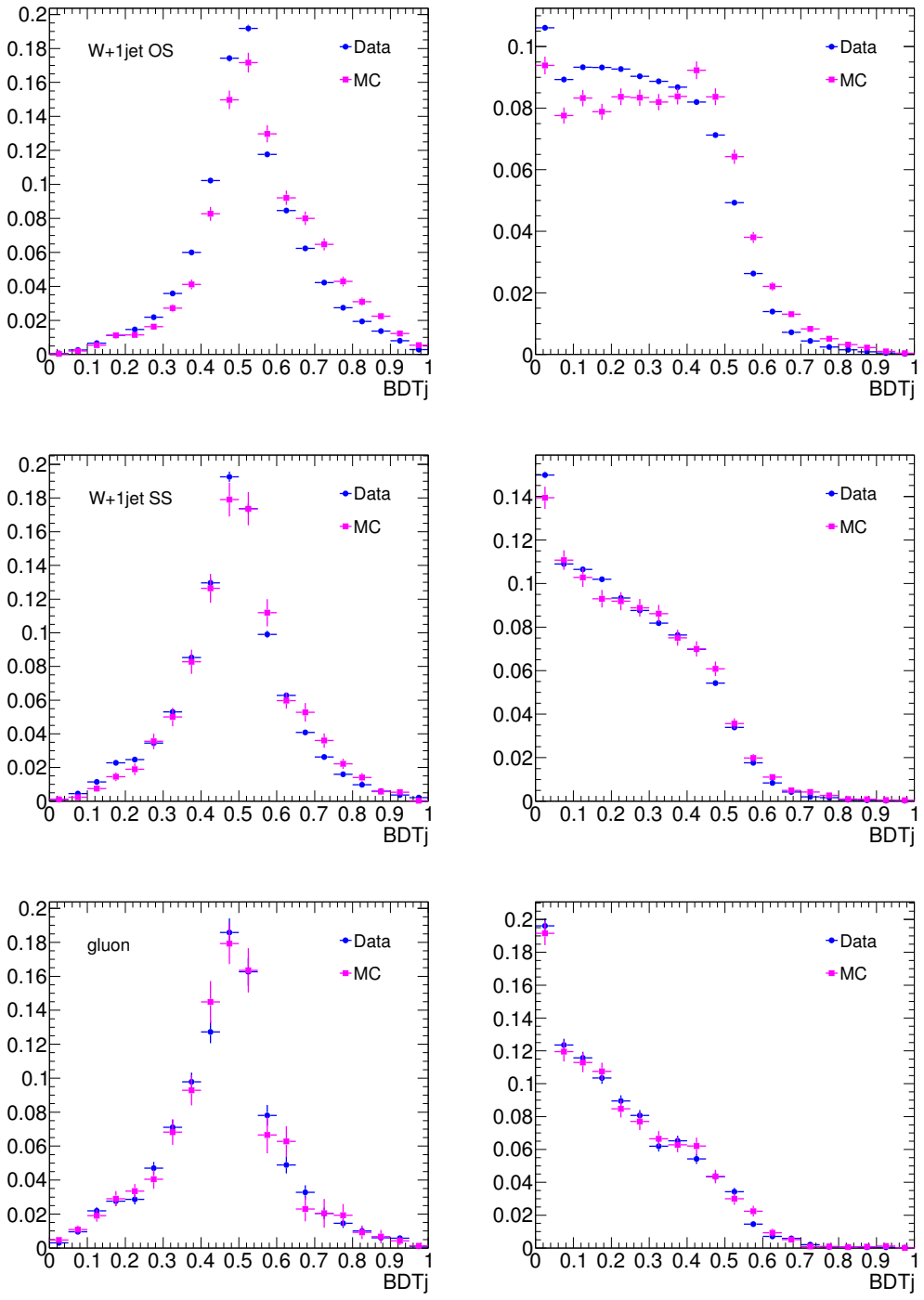
$$p = \frac{R \cdot N_{evt}(\text{SS1}) - N_{evt}(\text{OS1})}{N_{evt}(\text{OS1})(R-1)}. \tag{B.7}$$

	τ_1		τ_3	
p_T (GeV)	MC	Data	MC	Data
20-25	17%	32%	24%	38%
25-35	11%	24%	18%	30%
35-50	0%	16%	12%	22%
50-100	1.2%	13%	7.7%	16%
20-100	10%	25%	17%	29%

Table B.1: *Percentage of gluon jets misidentified as τ candidates in OS W + 1 jet simulated (MC) and observed (data) events for τ_1 and τ_3 depending on the p_T of the reconstructed τ candidate.*

Table B.1 shows the gluon content (as % of W + 1 jet OS events) observed in data and MC for different ranges of the p_T of the τ candidate. With these numbers we can derive the shapes

Figure B.1: *BDT_j distributions for τ candidates from OS W + 1 jet (top row), SS W + 1 jet (middle row) and gluon jets (bottowm row) samples. The left column is for τ_1 and right column for τ_3.*



of the OS_q and OS_q BDT_j distributions for each p_T range by subtracting the gluon component from the W + 1 jet distributions:

$$OS(SS)_q = OS1(SS1) - g \cdot \text{Gluons}, \tag{B.8}$$

where g is calculated from:

$$g \cdot N_{evt}(\text{Gluons}) = p \cdot N_{evt}(OS1). \tag{B.9}$$

Figures B.2 and B.3 compare the MC and data normalized BDT_j distributions for OS_q, SS_q and gluons. SS_q and gluon distributions match reasonably well for both τ_1 and τ_3. The OS_q data and MC distributions have similar shapes but MC ones have significantly higher average values. In conclusion, most of the difference between W + 1 jet data and MC comes from larger gluon content in the data but some is also due to MC not producing accurate OS_q distributions.

These shape discrepancies between the data and MC W + 1 jet BDT_j distributions reinforce the decision to extract the background distributions from data for the fitting method cross section measurement developed in chapter 5.

Figure B.2: *BDT_j normalized distributions for τ_1 candidates from OS_q, SS_q and gluon jets samples. BDT_j distributions are shown both for data and MC for several ranges on the p_T of the τ candidate.*

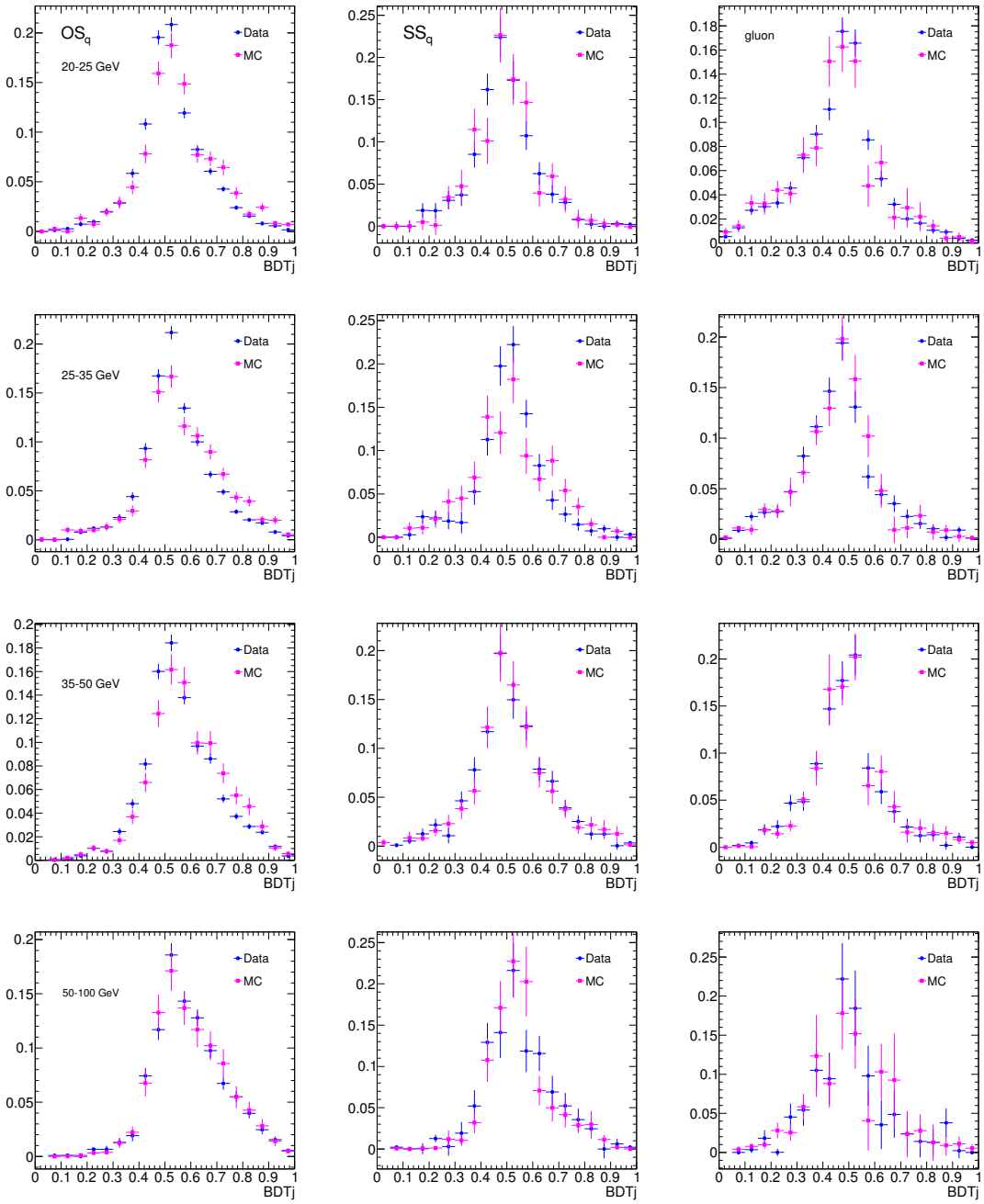
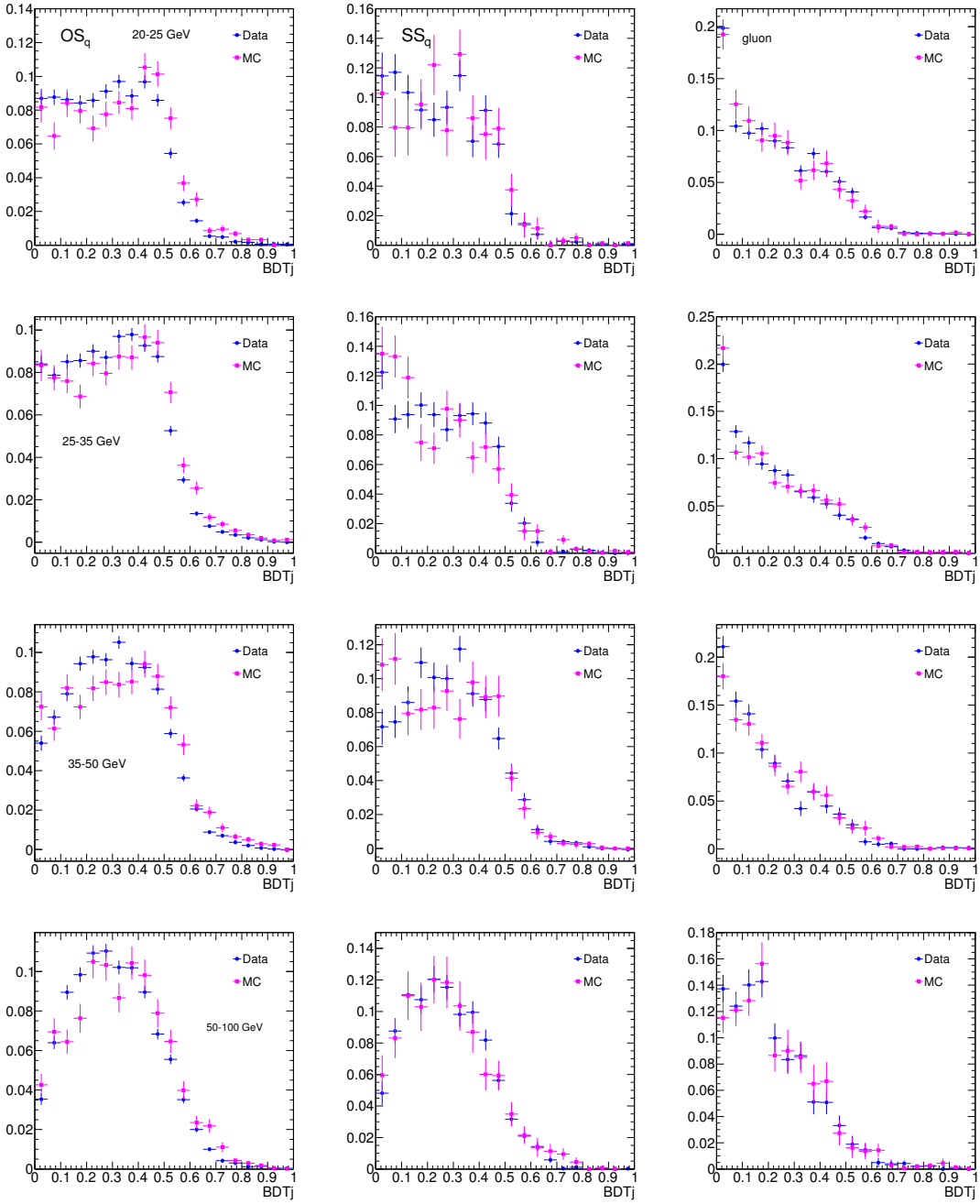


Figure B.3: *BDT_j normalized distributions for τ_3 candidates from OS_q, SS_q and gluon jets samples. BDT_j distributions are shown both for data and MC for several ranges on the p_T of the τ candidate.*



C

τ lepton ID systematic with the Fitting Method

As we saw in chapter 5, the measurement of the $t\bar{t}$ production cross section in proton-proton collisions at $\sqrt{s} = 7$ TeV ($\sigma_{t\bar{t}}$) using final states with one lepton (electron or muon) and one hadronically decaying τ lepton relies on extracting the number of real τ leptons in a sample of τ candidates by fitting the distribution of a Boosted Decision Tree output (BDT_j) trained to separate real taus from other type of jets misidentified as τ leptons. The fit uses background templates derived from data and signal templates derived from Monte Carlo.

The systematic uncertainty for extracting a τ signal due to the τ identification technique itself, is estimated by fitting BDT_j distributions from a $Z \to \tau^+\tau^-$ selection from data with background templates derived from a W + jets data selection and signal templates derived from MC. The number of τ leptons obtained with this fit is then compared with the number predicted from MC relying on the Z production cross section measured by ATLAS [156].

The $Z \to \tau^+\tau^-$ sample is selected from data with the same trigger, event cleaning and object identification criteria as in the data set used for measuring the $t\bar{t} \to \ell + \tau$ cross section, described in section 3.2, but a different kinematic selection criteria to enhance the number and purity of Z events. These requirements are:

- events must pass a single lepton trigger: a single-muon trigger with a p_{T} threshold of 18 GeV or a single-electron trigger with a p_{T} threshold of 20(22) GeV.
- a primary vertex with at least five tracks must be present in the event
- events are discarded if any jet with $p_{\mathrm{T}} > 20$ GeV fails jet quality selections designed to reject jets arising from out-of-time activity or calorimeter noise
- at least one τ candidate

Jet Composition of τ candidates in the $Z \to \tau\tau$ sample

τ_1	Events	gluon	quark	b	τ	e	μ
$t\bar{t}$	46	3%	45%	13%	38%	1%	0%
Zee	4871	6%	33%	0%	0%	61%	0%
$Z\mu\mu$	986	17%	71%	0%	0%	0%	12%
$Z\tau\tau$	10630	1%	4%	0%	95%	0%	0%
W+ jets	10306	13%	87%	0%	0%	0%	0%
Single Top	37	5%	50%	22%	22%	1%	0%
Diboson	118	2%	69%	1%	26%	2%	0%
Total MC	26995	7%	44%	0%	38%	11%	0%
τ_3	Events	gluon	quark	b	τ	e	μ
$t\bar{t}$	92	9%	60%	24%	7%	0%	0%
Zee	4966	23%	71%	0%	0%	6%	0%
$Z\mu\mu$	2210	28%	72%	0%	0%	0%	0%
$Z\tau\tau$	4876	8%	20%	0%	72%	0%	0%
W+ jets	24997	22%	78%	0%	0%	0%	0%
Single Top	95	4%	67%	26%	3%	0%	0%
Diboson	190	8%	87%	1%	4%	0%	0%
Total MC	37425	20%	69%	0%	9%	1%	0%

Table C.1: *Composition of all OS τ candidates in MC events, after requiring the $Z \to \tau\tau$ selections.*

- no more than one jet not overlapping with a tau candidate
- one and only one isolated lepton. An electron for the $e + \tau$ channel and a muon for the $\mu + \tau$ channel
- $M_{\mathrm{T}} < 20$ GeV to reduce the W+ jets background in the $Z \to \tau^+\tau^-$ samples
- All data samples are split into two sub-samples, one with τ and lepton having the opposite sign charge (OS), and the other with τ and lepton having the same sign charge (SS). The τ leptons contribute only to the OS samples.

With these selections we obtain a sample where the τ candidates are a mixture of mainly real taus, gluon and light-quark jets, as can be seen in Table C.1.

We can, therefore, fit the $Z \to \tau^+\tau^-$ BDT_j OS data distribution with light-quark and gluon jet background templates derived from the W+1 jet and W+2 jets data samples as described in 5.4, and a signal template derived from MC requiring the reconstructed τ candidate to be matched with a true τ. The BDT_j distributions for τ_1 and τ_3 are fitted separately since their shapes are considerably different.

Figure C.1 shows the fit results in the two p_T bins ($p_{\mathrm{T}} < 35$ GeV and $p_{\mathrm{T}} > 35$ GeV). The same fits, using the combined p_{T} fitting technique described in 5.5.2.1, are shown in Figure C.2. Table C.2 shows the number of τ leptons obtained by both fits with the corresponding χ^2/ndf and the number of MC expected τ leptons.

The statistical uncertainties on the number of τ leptons from fits to OS BDT_j distributions in two independent p_{T} bins fits are 3.2% for τ_1 and 7.9% for τ_3, slightly higher than those obtained

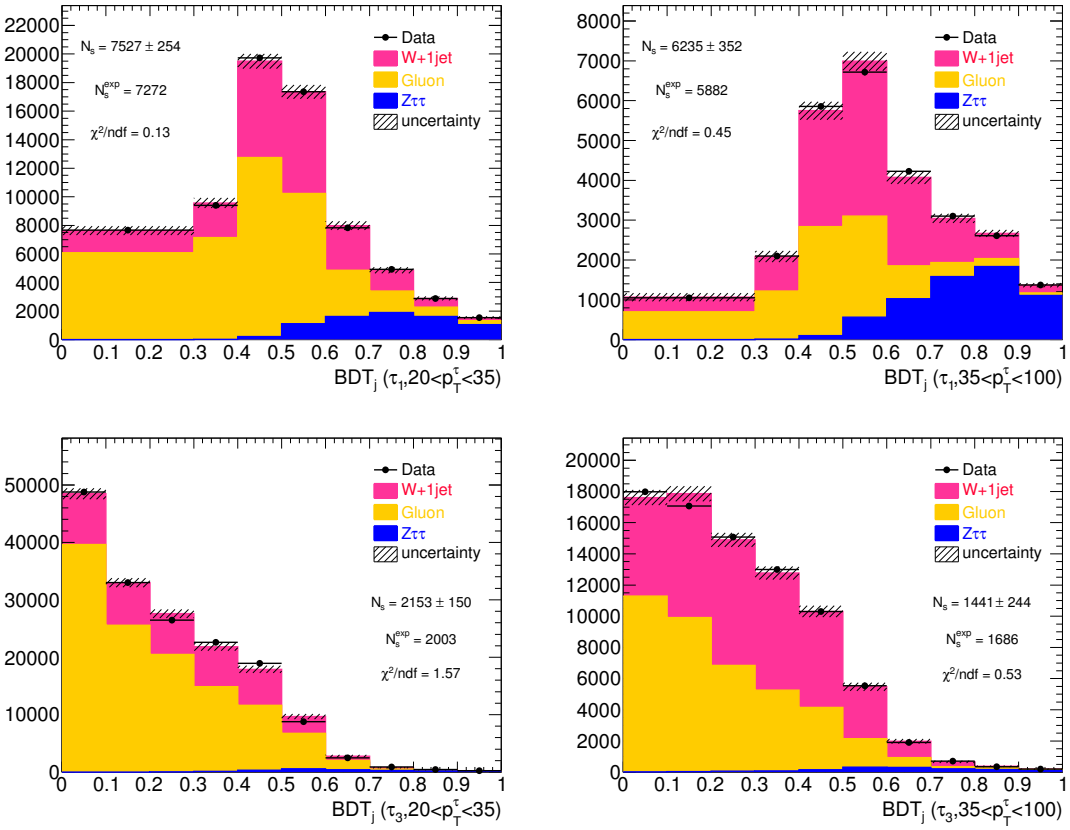


Figure C.1: *BDT_j OS distributions for τ_1 (top) and τ_3 (bottom) in two p_T bins. The points are data, solid orange histogram is the amount of gluon jet background, solid pink histogram is the extra amount of W + 1 jet background. The solid blue histogram is the amount of τ leptons from Z decays.*

	τ_1				
	$20 < p_T < 35$	$35 < p_T < 100$	Merged p_T bins	MC	Combined Fits
$Z \to \tau^+\tau^-$	7527 ± 254	6235 ± 353	13763 ± 435	13155 ± 86	14562 ± 383
χ^2/ndf	0.13	0.44		-	0.96
Fit Uncertainty	-	-	3.2%	-	2.6%
	τ_3				
	$20 < p_T < 35$	$35 < p_T < 100$	Merged p_T bins	MC	Combined Fits
$Z \to \tau^+\tau^-$	2153 ± 150	1442 ± 244	3596 ± 287	3689 ± 46	3587 ± 323
χ^2/ndf	1.57	0.53		-	0.99
Fit Uncertainty	-	-	7.9%	-	9.0%

Table C.2: *$Z \to \tau^+\tau^-$ fit results. The first row is the number of $Z \to \tau^+\tau^-$ events obtained from fitting the BDT_j distributions and the expected number of events from MC. The second row shows the reduced χ^2 of the fits and the third one the data/MC deviation.*

from fits with OS-SS BDT_j distributions (3.1% for τ_1 and 5.9% for τ_3) [104]. This results are consistent with those obtained with the full 5 fb^{-1} 2011 dataset (2.4% for τ_1 and 5.1% for τ_3) [155]. The statistical uncertainties of the combined p_T fits are 2.6% for τ_1 and 9.0% for τ_3.

The total uncertainty on τ acceptance, or tau identification uncertainty, includes the fit statistical uncertainty, the Z cross section uncertainty (2.1%, excluding luminosity) and the jet energy scale uncertainty introduced by the jet selection and M_T cut (3.3%). All these effects considered, the total tau identification uncertainty is 5.0% (4.7%) for τ_1 and 8.8% (9.8%) for τ_3.

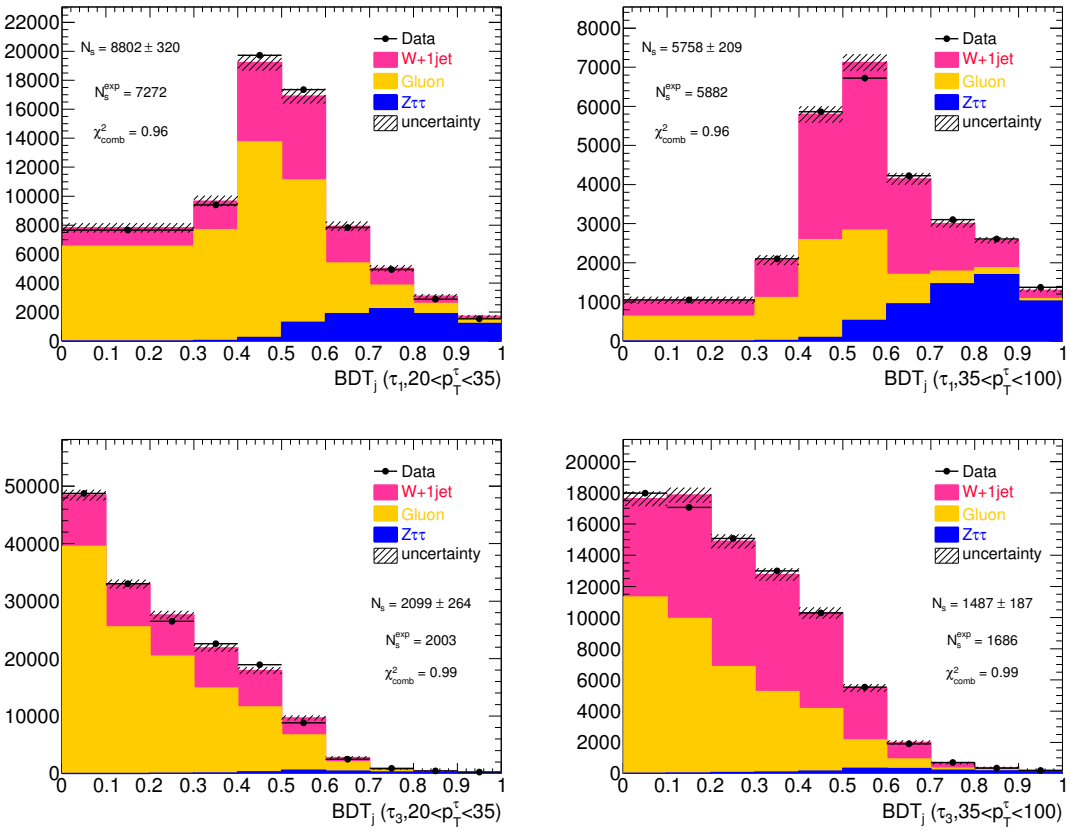


Figure C.2: *Fits of $Z \rightarrow \tau^+\tau^-$ BDT_j OS distributions from data for τ_1 (top) and τ_3 (bottom) using the combined p_T bins technique. The points are data, solid orange histogram is the amount of gluon jet background, solid pink histogram is the extra amount of W + 1 jet background. The solid blue histogram is the amount of τ leptons from Z decays.*

Acknowledgement

I would like to thank all the people who have made the work presented in this thesis possible.

First and foremost I want to thank my supervisor, Susana Cabrera, for her guidance and advice throughout the years. She has supported me academically and also emotionally through the rough road to finish this thesis, giving me the moral support and the freedom I needed to go on. It has been an honour to be her first student.

I want to give special thanks to Serban Protopopescu for his valuable leadership and contagious enthusiasm. You have been an inspiration to me and it has been a privilege working with you.

I would like to acknowledge the support of my collaborators Eva Valladolid, Yuta Takahashi and Jacob Searcy. I have learnt a lot and it has been a pleasure working with you all.

The members of the aTOPe group at the Instituto de Física Corpuscular have contributed immensely both professionally and personally. I want to thank Carmen, Salva, María Jose, Santi, Pepe, Farida, Juan and Vasia for their good advice and suggestions.

I can not forget my fellow PhD students, María, Miguel, Emma, Regina, Vicente, Adrián, Elena, Vicki, Sebas and Patricia. You are more than colleagues, you are my friends. Thank you for such wonderful years, for taking me through the difficult times, for making me laugh. Elena Oliver, you have been like a sister to me and my personal user support, without you this analysis would not have been completed.

Thanks to my non physicist friends, for pretending you were interested in my work and your continuous efforts to show me there is a life beyond PhD.

Last, but not least, I would like to thank my amazing family. I would not have made it this far without them. Dad, I know you have been with me every step of the way, I love you.

Bibliography

[1] D.H. Perkins. Introduction to high energy physics. *Cambridge University Press*, 2000.

[2] M. Veltman Hooft. *Nuc. Phys. B.*, 44:189, 1972.

[3] S.L. Glashow. Partial symmetries of weak interactions. *Nuc. Phys.*, 22:579, 1961.

[4] S. Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.

[5] Salam. In Nils Svartholm, editor, *Proceedings of the Eighth Nobel Symposium*, pages 367–377. Almquist & Wiksell.

[6] W. James Stirling R. Keith Ellis and Bryan R. Webber. *QCD and Collider Physics*. Cambridge University Press, 1996.

[7] Gene Mosca Paul A. Tipler. *Modern Physics*. 2003.

[8] Particle Data Group website. http://pdg.lbl.gov/.

[9] P.W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys. Lett.*, 12:132, 1964.

[10] P.W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508, 1964.

[11] P.W. Higgs. Spontaneous symmetry breakdown without massless bosons. *Phys. Rev.*, 145:1156, 1966.

[12] R. Brout F. Englert. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321, 1964.

[13] The ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1–29, 2012.

[14] The CMS Collaboration. Observation of a new boson at a mass of 125 GeVwith the CMS experiment at the LHC. *Phys. Lett. B*, 716:30, 2012.

[15] The ATLAS collaboration. Combined measurements of the mass and signal strength of the Higgs-like boson with the ATLAS detector using up to 25 fb^{-1} of proton-proton collision data. *ATLAS-CONF-2013-014*.

[16] The CMS collaboration. Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s}$ = 7 and 8 TeV. *High Energy Phys.*, 06:081, 2013.

[17] The ATLAS collaboration. Evidence for the spin-0 nature of the Higgs boson using ATLAS data. *Physics Letter*, B726:120–144, 2013.

[18] The ATLAS collaboration. Combined coupling measurements of the Higgs-like boson with the ATLAS detector using up to 25 fb^{-1} of proton-proton collision data. *ATLAS-CONF-2013-034*.

[19] Stephen P. Martin. A supersymmetry primer. *arXiv:hep-ph/9709356*.

[20] H.P. Nilles. Supersymmetry, supergravity and particle physics. *Phys. Rep.*, 110:1, 1984.

[21] G.L. Kane H.E. Haber. The search for supersymmetry: Probing physics beyond the standard model. *Phys. Rep.*, 117:75, 1985.

[22] Summary of SUSY searches with ATLAS. https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SUSY/ATLAS_SUSY_Summary/ATLAS_SUSY_Summary.pdf.

[23] CDF website. http://www-cdf.fnal.gov/.

[24] DØwebsite. http://www-d0.fnal.gov/.

[25] F. Abe *et al.* Observation of top quark production in $p\bar{p}$ collisions with the collider detector at fermilab. *Phys. Rev. Lett.*, 74(14):2626–2631, Apr 1995.

[26] D. Chakraborty, J. Konigsberg, D.L. Rainwater. Review of top quark physics. *Ann.Rev.Nucl.Part.Sci.*, 53, 2003.

[27] MC datasets relevant for 2010 analysis. https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TopMC2010.

[28] MC datasets relevant for 2011 analysis. https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TopMC10For2011Data.

[29] M. Aliev et al., HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR, Comput. Phys. Commun. **182** (2011) 1034, arXiv:1007.1327.

[30] A. Mitov M. Czakon, P. Fiedler. The total top quark pair production cross-section at hadron colliders through $o(\alpha_S^4)$. *Phys. Rev. Lett.*, 110, 2013.

[31] F. Dliot, Y. Peters, V. Sorin. Top quark Physics at the Tevatron. *arXiv:1302.3628*, 2013.

[32] Carlo Battilana, The ATLAS and CMS Collaborations. Top quark Physics at the LHC. *arXiv:1309.5307*, 2013.

[33] Frank-Peter Schilling. Top Quark Physics at the LHC: A Review of the First Two Years. *Int. J. Mod. Phys.*, A27, 2012.

[34] Combination of CDF and DO results on the mass of the top quark using up to 8.7 fb^{-1} at the Tevatron. *arXiv:hep-ex/1305.3929*, 2013.

[35] The ATLAS and CMS Collaborations. Combination of ATLAS and CMS results on the mass of the top quark using up to 4.9 fb^{-1} of $\sqrt{s}$ = 7 TeV LHC data. (ATLAS-CONF-2013-102, CMS PAS TOP-13-005), 2013.

[36] Combination of measurements of the top-quark pair production cross section from the Tevatron Collider. *arXiv:hep-ex/1309.7570*, 2013.

[37] Combination of ATLAS and CMS top-quark pair cross-section measurements using proton-proton collisions at $\sqrt{s}$ = 7 TeV. *ATLAS-CONF-2012-134, CMS PAS TOP-12-003*, 2012.

[38] Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s}$ = 8 TeV. *CMS-TOP-12-007, CERN-PH-EP-2013-234, arXiv:1312.7582*, 2013.

[39] Measurement of the top quark pair production cross section in the single-lepton channel with ATLAS in proton-proton collisions at 8 TeV using kinematic fits with b-tagging. *ATLAS-CONF-2012-149*, 2012.

[40] ATLAS Collaboration, *Measurement of the top quark pair cross-section Measurements in ATLAS in the Single lepton channel using b-tagging*, ATLAS-CONF-2011-035.

[41] ATLAS Collaboration, *Top Quark Pair Production Cross-section Measurement in ATLAS in the Single Lepton+Jets Channel without b-tagging*, ATLAS-CONF-2011-023.

[42] The ATLAS Collaboration, *Measurement of the top quark pair production cross-section with ATLAS in pp collisions at* $\sqrt{s}$ = 7 *TeV in dilepton final states*, ATLAS-CONF-2011-034.

[43] The ATLAS Collaboration, *Measurement of the cross section for top-quark pair production in pp collisions at* $\sqrt{s}$ = 7 *TeV with the ATLAS detector using final states with two high-*p_T *leptons*, JHEP 1205 (2012) 059.

[44] The ATLAS collaboration, *Search for ttbar production in the all-hadronic channel in ATLAS with* $\sqrt{s}$ = 7 *TeV data*, ATLAS-CONF-2011-066, 2011.

[45] CMS collaboration. Measurement of the $t\bar{t}$ production Cross Section in pp collsions at $\sqrt{s}$ = 7 TeV in Lepton + Jets events using b-quark jet identification. *Phys. Rev. D*, 84:092004, 2011.

[46] CMS collaboration. Measurement of the $t\bar{t}$ production cross section in pp collsions at $\sqrt{s}$ = 7 TeV with lepton + jets final states. *Phys. Lett. B*, 720:83, 2013.

[47] CMS collaboration. Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collsions at $\sqrt{s}$ = 7 TeV. *JHEP*, 11:067, 2012.

[48] The D0 Collaboration. First measurement of the forward-backward charge asymmetry in top quark pair production. *Phys. Rev. Letter*, 100:142002, 2008.

[49] The CDF Collaboration. Measurement of the top quark forward-backward production asymmetry and its dependence on event kinematic properties. *Phys. Rev. D*, 87, 2013.

[50] The CMS Collaboration. Measurements of the $t\bar{t}$ charge asymmetry using the dilepton decay channel in pp collisions at $\sqrt{s}$ = 7 TeV. *arXiv:1402.3803*, 2014.

[51] The ATLAS Collaboration. Measurement of the top quark pair production charge asymmetry in proton-proton collisions at 7 TeV using the ATLAS detector. *JHEP*, 02:107, 2014.

[52] Werner Bernreuther, Zong-Guo Si. Top quark and leptonic charge asymmetries for the Tevatron and LHC. *Phys. Rev. D.*, 86, 2012.

[53] The ATLAS Collaboration. Measurement of top quark polarization in top-antitop events from proton-proton collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector. *arXiv:1307.6511*, 2013.

[54] The CMS Collaboration. Measurements of $t\bar{t}$ spin correlations and top-quark polarization using dilepton final states in pp collisions at 7 TeV. *arXiv:1311.3924*, 2013.

[55] The ATLAS Collaboration. Observation of Spin Correlation in $t\bar{t}$ Events from pp Collisions at $\sqrt{s}$ = 7 TeV using the ATLAS Detector. *Phys. Rev. Letter*, 108:212001, 2012.

[56] The ATLAS Collaboration. A search for flavour changing neutral currents in top-quark decays in pp collision data collected with the ATLAS detector at $\sqrt{s}$ = 7 TeV. *JHEP*, 09:139, 2012.

[57] The CMS Collaboration. Search for flavor-changing neutral currents in top-quark decays to Zq in pp collisions at $\sqrt{s}$ = 8 TeV. *arXiv:1312.4194*, 2013.

[58] The DØ collaboration. Search for charged Higgs bosons in top quark decays. *Phys. Lett. B*, 682:278, 2009.

[59] The ATLAS collaboration. Search for charged Higgs bosons decaying via $H^{\pm} \rightarrow \tau\nu$ in $t\bar{t}$ events in pp collsion data at $\sqrt{s}$ = 7 TeV with the ATLAS detector. *JHEP*, 1206:039, 2012.

[60] The CMS collaboration. Search for a light charged Higgs boson in top quark decays in pp collisions at $\sqrt{s}$ = 7 TeV. *JHEP*, 07:143, 2012.

[61] The ATLAS collaboration. Search for charged Higgs bosons in the τ + jets final states with pp collsion data at $\sqrt{s}$ = 8 TeV with the ATLAS detector. *ATLAS-CONF-2013-090*, 2013.

[62] et. al. M. Carena. Suggestions for benchmark scenarios for mssm higgs boson searches at hadron colliders. *Eur.Phys.J.*, C26:601–607, 2003.

[63] V. M. Abazov et al., The DØ collaboration. Measurement of $t\bar{t}$ production cross section in the lepton + tau + b-jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ channel using $1.2 fb^{-1}$ of run IIb data. (5607-CONF), 2008.

[64] The DØ collaboration. Combination of $t\bar{t}$ cross section measurements and constraints on the mass of the top quark and it's decays into charged Higgs bosons. *Phys. Rev. D*, 80:071102(R), 2009.

[65] T. Aaltonen, CDF collaboration. Measurement of the top pair cross section in dileptonic channels with a hadronic tau and branching ratio $t \to \tau\nu b$, with 9.0 fb^{-1}. *arXiv:1402.6728*, 2012.

[66] CMS collaboration. Measurement of the top quark pair production cross section in pp collsions at $\sqrt{s}$ = 7 TeV in the dilepton final states containing a tau. *Phys. Rev. D*, 85:112007, 2012.

[67] ATLAS collaboration, *Measurement of the top quark pair production cross section with ATLAS in pp collisions at* $\sqrt{s}$ = 7 *TeV using final states with an electron or a muon and a hadronically decaying* τ *lepton*, CERN-PH-EP-2012-102, arXiv:1205.2067, Phys. Lett. B 717 (2012) 89-108.

[68] L. Evans and P. Bryant (editors). LHC Machine. 2008. JINST 3 S08001.

[69] CERN website. http://www.cern.ch.

[70] http://press-archived.web.cern.ch/press-archived/PressReleases/Releases2008/PR14.08E.html.

[71] Mike Lamont. Status of the LHC. *Journal of Physics*, Conference Series 455, 2013.

[72] Mirko Pojer. LHC Machine: Status and Plan. *EPJ Web Conf.*, 60, 2013.

[73] The ATLAS Collaboration. Physics at a High-Luminosity LHC with ATLAS. *ATL-PHYS-PUB-2013-007*.

[74] ATLAS experiment website. http://atlas.web.cern.ch/Atlas/Collaboration/.

[75] The ATLAS Collaboration, G. Aad *et al.* The ATLAS Experiment at the CERN Large Hadron Collider. 2008. JINST 3 S08003.

[76] The CMS Collaboration, S. Chatrchyan *et al.* The CMS experiment at the CERN Large Hadron Collider. 2008. JINST 3 S08004.

[77] The LHCb Collaboration, A. Augusto Alves Jr *et al.* The LHCb Detector at the Large Hadron Collider. 2008. JINST 3 S08005.

[78] The ALICE Collaboration, K. Aamodt *et al.* The ALICE experiment at the CERN Large Hadron Collider. 2008. JINST 3 S08002.

[79] The TOTEM Collaboration, G. Anelli *et al.* The TOTEM Experiment at the CERN Large Hadron Collider. 2008. JINST 3 S08007.

[80] The LHCf Collaboration, O. Adriani *et al.* The LHCf detector at the CERN Large Hadron Collider. 2008. JINST 3 S08006.

[81] The ATLAS Collaboration. ATLAS detector and physics performance: Technical Design Report,1. Technical report, 1999.

[82] ATLAS luminosity public results. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults.

[83] The ATLAS Collaboration. Updated Luminosity Determination in pp Collisions at $\sqrt{s}$ = 7 TeV. (ATLAS-CONF-2011-011), 2011.

[84] The ATLAS Collaboration. Luminosity Determination in pp Collisions at $\sqrt{s}$ = 7 TeV using the ATLAS Detector in 2011. (ATLAS-CONF-2011-116), August 2011.

[85] The ATLAS Collaboration. Improved Luminosity Determination in pp Collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector at the LHC. *Eur. Phys. Journal*, C73, 2013.

[86] Ian Foster and Carl Kesselman. The Grid: Blueprint for a New Computing Infrastructure. *Morgan Kaufmann*, 2003.

[87] WLCG. http://wlcg.web.cern.ch.

[88] Christoph Eck et al. LHC computing Grid: Technical Design Report. 2005.

[89] The ATLAS collaboration. ATLAS computing: Technical Design Report. 2005.

[90] I Ueda. ATLAS Distributed Computing Operations in the First Two Years of Data Taking. Technical Report ATL-SOFT-PROC-2012-003, 2012.

[91] Athena Analysis Framework. http://atlas-computing.web.cern.ch/atlas-computing/packages/athenaCore/athenaCore.php.

[92] ROOT: A Data Analysis Framework. http://root.cern.ch/drupal/.

[93] Adelman, J. et al. Technical aspects of the first top pair analyses. Technical Report ATL-PHYS-INT-2010-140, December 2010.

[94] GEANT4 collaboration, S. Agostinelli et. al., "A simulation toolkit", Nuc. Instrum. and Meth. A506 (2003) 250.

[95] The ATLAS collaboration, "The ATLAS Simulation Infraestructure", Eur. Phys. J. C70 (2010) 823.

[96] S. Frixione and B.R. Webber. Matching NLO QCD computations and parton shower simulations. *JHEP*, 0206:029, 2002.

[97] S. Frixione, P. Nason and B.R. Webber. Matching NLO QCD and parton showers in heavy flavor production. *JHEP*, 0308:007, 2003.

[98] S. Frixione, E. Laenen and P. Motylinski. Single-top production in MC@NLO. *JHEP*, 0603:092, 2006.

[99] P.M. Nadolsky et al. Implications of CTEQ global analysis for collider observables. *Phys.Rev.D*, 78:013004, 2008.

[100] C.D. White, et. al. *JHEP*, 11:74, 2009.

[101] S. Frixione, et. al. *JHEP*, 07:29, 2008.

[102] M.L. Mangano, M. Moretti, H. Lai, P. Nadolsky and A.D. Polosa, ALPGEN, a generator for hard multiparton processes in hadronic collisions, JHEP 07 (2003) 001.

[103] J. Alwall et al. Comparative studies of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J.*, C53:473, 2008.

[104] J. Godfrey et al. Measurement of hadronic tau identification using Boosted Decision Trees and Z events. (ATL-COM-PHYS-2011-1686), December 2011.

[105] G. Corcella et al. Herwig 6.5: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes). *JHEP*, 0101:010, 2001.

[106] G. Corcella et al. Herwig 6.5 release notes. *CAVENDISH-HEP-02-17,arXiv:hep-ph/0210213*, 2002.

[107] J. Butterworth et al. Multiparton interactions in photoproduction at hera. *Zeit. fur Phys*, C72:637, 1996.

[108] N.Davidson et al. Universal Interface of TAUOLA Technical and Physics Documentation. *arXiv:1002.0543*, 10, 2009.

[109] The ATLAS collaboration, ATLAS tunes of PYTHIA 6 and PYTHIA 8 for MC11, ATL-PHYS-PUB-2011-009.

[110] TopCommonObjects Selection for 2010 analysis. https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TopCommonObjects2010.

[111] TopCommonObjects Selection for 2011 analysis with release 16. https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TopCommonObjects2011rel16.

[112] The ATLAS collaboration, Lepton trigger and identification for the Winter 2011 top quark analyses, ATL-COM-PHYS-2011-123.

[113] Matteo Cacciari and Gavin P. Salam. Dispelling the N3 myth for the kt jet-finder. *Physics Letters B*, 641(1):57 – 61, 2006.

[114] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-kT jet clustering algorithm. *JHEP*, 04:063, 2008.

[115] ATLAS Collaboration. Measurement of inclusive jet and dijet cross sections in proton-proton collisions at 7 TeV centre-of-mass energy with the ATLAS detector. *Eur.Phys.J.C71:1512*, 2011.

[116] S. Adomeit et. al. Jets, Missing Transverse Energy and Taus for Top Physics Analyses in Release 16 with the 2010 Dataset. (ATL-COM-PHYS-2011-132), 2011.

[117] Nugent, I. ; Rosbach, K. ; Zhu, C. Missing Transverse Energy for Top Analyses in Release 16.6.5.5.1 with the 2011 dataset. (ATL-COM-PHYS-2011-1565), 2011.

[118] BTagging For Early Data. https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/BTaggingForEarlyData.

[119] The ATLAS Collaboration. Performance of the ATLAS Secondary Vertex b-tagging Algorithm in 7 TeV Collision Data. *ATLAS-CONF-2010-042*.

[120] The ATLAS Collaboration. Commisioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data. *ATL-CONF-2011-102*.

[121] The ATLAS Collaboration. Determination of the tau energy scale and the associated systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector at the LHC in 2011. (ATLAS-CONF-2012-054), 2012.

[122] The ATLAS Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons*,ATL-CONF-2011-077.

[123] The ATLAS Collaboration. Performance of the Reconstruction and Identification of Hadronic Tau Decays with ATLAS. (ATLAS-CONF-2011-152), Nov 2011.

[124] E. Coniavitis et. al., on behalf of the Tau Combined Performance Group, *Reconstruction, Energy Calibration, and Identification of Hadronic Tau Decays for Winter 2011*,ATL-COM-PHYS-2011-217.

[125] Philip Bechtle, Bjorn Gosdzik, Stan Lai. Cut based reconstruction and identification of Hadronic tau Decays. (ATL-PHYS-INT-2009-082, ATL-COM-PHYS-2009-304, ATL-PHYS-PUB-2010-001).

[126] N. Dawe, D. O'Neil and S. Protopopescu. Using Boosted Decision Trees for Hadronic Tau Identification. ATL-PHYS-INT-2011-004, ATL-COM-PHYS-2010-603.

[127] B.P. Roe et al. Boosted decision trees, an alternative to artificial neural networks. *Nucl.Instrum.Meth.A*, 543:577–584, 2005.

[128] Y. Freund. Boosting a weak learning algorithm by majority, in COLT 90: Proceedings of the third annual workshop on Computational learning theory. 1990.

[129] V. Abazov et al. The DØ Collaboration. Observation of single top-quark production. *Phys. Rev. Lett.*, 103(9):092001, 2009.

[130] The ATLAS Collaboration. Data-Quality Requirements and Event Cleaning for Jets and Missing Transverse Energy Reconstruction with the ATLAS Detector in Proton-Proton Collisions at a Center-of-Mass Energy of $\sqrt{s} = 7$ TeV. Technical Report ATLAS-CONF-2010-038, Jul 2010.

[131] The ATLAS Collaboration. ATLAS Sensitivity of Tau Identification Techniques in $t\bar{t}$ events. (ATLAS-COM-CONF-2011-032), 2011.

[132] The ATLAS Collaboration. Determination of the ATLAS jet energy measurement uncertainty using photon-jet events in proton-proton collisions at $\sqrt{s} = 7$ TeV. (ATLAS-CONF-2011-031), 2011.

[133] S. Bedekian, C. Bernius, S. Cabrera Urbn, S. Demers, S. Protopopescu et. al. , *Measurement of the top quark pair production cross section in pp collisions at* $\sqrt{s} = 7$ *TeV in the* $\mu + \tau$ *Channel with ATLAS*, ATLAS-COM-PHYS-2011-943.

[134] ATLAS collaboration, *Measurement of the top quark pair production cross section in pp collisions at* $\sqrt{s} = 7$ *TeV in* $\mu + \tau$ *final states with ATLAS*, ATLAS-CONF-2011-119.

[135] S. Bedekian et. al., *Measurement of the top quark pair production cross section in pp collisions at* $\sqrt{s} = 7$ *TeV in the* $\ell + \tau$ *Channel with ATLAS*, ATL-COM-PHYS-2011-1696.

[136] Yuta Takahashi. *Measurement of the top-quark pair production cross-section in pp collisions at* $\sqrt{s} = 7$ TeV *using final states with an electron or a muon and a hadronically decaying* τ*-lepton*. PhD thesis, Nagoya University, 2012, CERN-THESIS-2012-252.

[137] Jacob Searcy. *Measurement of the Top quark production cross section in pp collisions at* $\sqrt{s} = 7$ TeV *in the* $\ell + \tau$ *channel with ATLAS*. PhD thesis, University of Oregon, 2012, CERN-THESIS-2012-190.

[138] K. Becker et al., "Mis-identified lepton backgrounds in top quark pair production studies for EPS 2011 analyses", ATL-COM-PHYS-2011-768, 2011.

[139] F. James and M. Roos. Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Comput.Phys.Commun.*, 10, 1975.

[140] P. Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

[141] S. Frixione, P. Nason, and C. Oleari. Matching NLO QCD computations with Parton Shower simulations: the Powheg method. *JHEP*, 11:070, 2007.

[142] S. Alioli, P. Nason, C.Oleari, and E. Re "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", JHEP 06 (2010) 043.

[143] J. Pumplin, D.R. Stump, J. Huston, H.L. Lai, P. Nadolsky and W.K. Tung "New Generation of Parton Distributions with Uncertainties from Global QCD Analysis", JHEP 07 (2002) 012, arXiv:hep-ph/0201195.

[144] CTEQ collaboration, H.L. Lai et al. "Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions", Eur. Phys. J. C12 (2000) 375, arXiv:hep-ph/9903282.

[145] A.D. Martin, R.G. Roberts, W.J. Stirling and R.S. Thorne, "Parton distributions: a new global analysis", Eur. Phys. J. C4 (1998) 463, "Parton distributions and the LHC: W and Z productions", Eur. Phys. J. C14 (2000) 133.

[146] B.P. Kersevan and E.Richter-Was "The Monte Carlo event generator AcerMC version 2.0 with interfaces to PYTHIA 6.2 and HERWIG 6.5", arXiv:hep-ph/0405247.

[147] ATLAS collaboration. Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data. *Eur. Phys. J.*, C72:1909, 2012.

[148] The ATLAS collaboration, Muon reconstruction efficiency in reprocessed 2010 LHC proton-proton collision data recorded with the ATLAS detector, ATLAS-CONF-2011-063.

[149] ATLAS Collaboration "Jet energy scale and its systematic uncertainty in proton-proton collisions at $\sqrt{s}$ = 7 TeV", ATLAS-CONF-2011-032.

[150] ATLAS Collaboration "Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s}$ = 7 TeV", arXiv:1112.6426, CERN-PH-EP-2011-191.

[151] L. Lyons, D. Gibaut, and P. Clifford. *Nuc. Instrum. Meth.*, A270:110, 1988.

[152] A. Valassi. *Nuc. Instrum. Meth.*, A500:391, 2003.

[153] S. Cabrera, S. Protopopescu, M.T. Perez, E. Valladolid, *Measurement of the top quark pair production cross section in pp collisions at $\sqrt{s}$ = 7 TeV in the $\ell + \tau$ channel with ATLAS*, ATL-COM-PHYS-2014-005.

[154] S. Cabrera, S. Protopopescu, *Measurement of the top branching ratios into channels with leptons and quarks using final states with $t\bar{t}$ pairs produced in pp collisions at $\sqrt{s}$ = 7 TeV*, ATL-COM-PHYS-2013-786.

[155] S. Protopopescu, S. Cabrera, M.T. Perez, E. Valladolid. Study of hadronic tau identification using W and Z events. (ATL-COM-PHYS-2012-1744), November 2012.

[156] The ATLAS Collaboration. Measurement of the $Z \to \tau\tau$ Cross Section with the ATLAS Detector. *Physics Rev. D*, 84, 2011.